



Dissertação

Mestrado em Finanças Empresariais

***O Uso de Redes Neurais Artificiais na  
Estimação do Preço das Habitações  
na Ilha do Sal, em Cabo Verde***

**Eunices Ester dos Santos Martins**

Leiria, *junho* de 2016





Dissertação

Mestrado em Finanças Empresariais

***O Uso de Redes Neurais Artificiais na  
Estimação do Preço das Habitações  
na Ilha do Sal, em Cabo Verde***

**Eunices Ester dos Santos Martins**

Dissertação de Mestrado realizada sob a orientação da Doutora Maria João da Silva Jorge,  
Professora da Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Leiria.

Leiria, *junho* de 2016



## AGRADECIMENTOS

Gostaria de agradecer a todos, que direta ou indiretamente, contribuíram para a realização desta dissertação. Concretamente agradeço:

- A toda a minha família, em especial, aos meus pais Manuel Martins e Maria Santos e aos meus irmãos pelo amor, carinho e todo o apoio necessário para alcançar este objetivo.
- À minha orientadora, a Professora Doutora Maria João da Silva Jorge, quer pelos esclarecimentos e sugestões na orientação do meu projeto, quer pela disponibilidade que manifestou ao longo da elaboração desta dissertação;
- À Djéniffa Semedo, Edilene Oliveira, Aristides Santos e Alexsander Barros, pela amizade e toda força que me deram nos momentos mais difíceis;
- Por fim agradeço à imobiliária por ter facultado os dados que tornaram possível a realização do projeto.

Muito obrigado a todos pois, sem a vossa ajuda seria impossível a realização desta dissertação.



## RESUMO

Esta dissertação tem como objetivo desenvolver um modelo de redes neuronais artificiais para aferir as características das habitações que mais influenciam o preço na Ilha do Sal, em Cabo Verde. Foram consideradas características como: área, número de quartos, existência de varandas, existência de terraços, número de casas de banho, localização do imóvel, número de andares e proximidade com instituições públicas. A amostra utilizada considerou 1092 habitações no período de 2009 a 2014. Para além da análise baseada no desenvolvimento do modelo de redes neuronais, efetuou-se a análise pela estimação do modelo dos preços hedónicos. Os resultados do modelo de redes neuronais artificiais permitiram verificar que o preço das habitações é fortemente influenciado pela área, e em seguida pela localização. A existência de características, tais como a proximidade com a câmara municipal e finanças e existência de varandas, são as variáveis que menos influenciam o preço das habitações na Ilha do Sal.

Os resultados da estimação com o modelo dos preços hedónicos indicam que o preço das habitações é fortemente influenciado por algumas variáveis representativas de características estruturais, localização e de vizinhança. Algumas dessas variáveis têm efeito estatisticamente significativo positivo no preço tais como, a localização do imóvel em Algodoeiro- Santa Maria, o número de quartos e a área. Outras variáveis têm efeito estatisticamente significativo negativo no preço, tais como a localização do imóvel no Bairro Novo e a proximidade com o hospital.

Os resultados mostram que comparativamente com o modelo de preços hedónicos, o modelo de redes neuronais artificiais representa uma melhor alternativa para a previsão dos preços das habitações na Ilha do Sal, isto considerando a comparação dos erros estimados entre os modelos e as medidas de desempenho comumente utilizadas.

Palavras-chave: redes neuronais artificiais, modelo de preços hedónicos, mercado imobiliário.



## **ABSTRACT**

This thesis aims to develop a model of artificial neural networks to assess the characteristics of the habitations that most influence the price on Sal Island in Cape Verde. Were considered characteristics such as: area, room numbers, existence of balconies and terraces, bathrooms numbers, property location, floor number and proximity to public institutions. The sample utilized considered 1092 habitations in the period from 2009 to 2014. In addition to the analysis based on the development of the neural network model, we estimated a model of hedonic prices. The results of the artificial neural network model allowed us to verify that the price of the habitations is strongly influenced by area, and then by location. The existence of characteristics such as terraces and bathrooms number are the variables that affects less the price of the habitations on Sal Island.

The results with the hedonic price model indicate that the price of the habitations is strongly influenced by some variables representative of structural features, location and neighbourhood. Some of these variables have positive effect on prices, such as the location of the property in Algodoeiro - Santa Maria, number of rooms and area. Other variables have a negative effect on prices, such as property located in Bairro Novo II and the proximity to the hospital.

Compared with the model hedonic price the results show that, the model of artificial neural networks is a better alternative to forecast the prices of habitations in Sal Island, considering the comparison of the estimated errors between models and performance measures commonly used.

Keywords: artificial neural networks, model of hedonic prices, real estate market.



## ÍNDICE

LISTA DE FIGURAS	xi
LISTA DE TABELAS	xiii
LISTA DE ANEXOS	xv
LISTA DE SIGLAS	xvii
1. INTRODUÇÃO	1
2. REVISÃO DE LITERATURA	5
2.1. Características do mercado imobiliário	5
2.2. Breve revisão dos métodos de avaliação imobiliário	6
2.2.1. Métodos de avaliação tradicionais	6
2.2.2. Métodos de avaliação avançados	8
2.2.2.1. Modelo de preço hedônico	8
2.2.2.2. Modelo de lógica vaga (Fuzzy Logic) e Modelos de análise espacial	12
2.3. Redes neuronais artificiais (RNA)	12
2.3.1. Perspetiva histórica do modelo das redes neuronais artificiais segundo (Fausett, 1994)	13
2.3.2. Características de redes neuronais artificiais	15
2.3.3. Ativação e função de ativação	16
2.3.4. Tipologia e treinamento de redes neuronais	18
2.3.5. Evidências empírica da utilização das Redes Neuronais no setor imobiliário	19
3. METODOLOGIA	29
3.1. Apresentação da Amostra	29
3.2. Identificação e Descrição das Variáveis	30
3.2.1. Variável dependente	30

3.2.2.	Variáveis independentes	31
3.3.	Forma funcional utilizado no MPH	36
3.4.	Justificação de utilização do modelo de RNA	38
3.5.	Processo de estimação do preço das habitações pela rede	39
3.6.	<i>Software</i> utilizado e medidas de desempenho da rede	39
4.	APRESENTAÇÃO E DISCUSSÃO DE RESULTADOS	43
4.1.	Estatística descritiva	43
4.2.	Estimação do preço das habitações através de RNA	45
4.3.	Estimação do preço das habitações através do MPH	52
4.4.	Comparação de desempenho do modelo de RNA e MPH	61
5.	CONCLUSÃO	65
6.	BILIOGRAFIA	67
7.	ANEXOS	73

## LISTA DE FIGURAS

Figura 1. Ativação de neurónio numa RNA.....	17
Figura 2. capacidades explicativas das variáveis independentes .....	49
Figura 3. Preço observado Vs. Preço estimado pela rede neuronal .....	50
Figura 4. Modelo do preço hedónico ajustado .....	56
Figura 5. Erro do modelo estimado.....	61
Figura 6. Comparação dos modelos com a variável área.....	63



## LISTA DE TABELAS

Tabela 1: Resumo da evidência empírica de alguns estudos que utilizam o MPH.	11
Tabela 2: Função Ativação da RNA .....	18
Tabela 3: Resumo de evidencia empírica que utiliza a metodologia das RNA no setor imobiliário .....	25
Tabela 4: Distribuição da amostra pelas cidades e zonas da Ilha do Sal .....	30
Tabela 5: Descrição e classificação dos atributos das habitações (RNA).....	31
Tabela 6: Descrição e classificação dos atributos das habitações (MPH) .....	33
Tabela 7: Resumo estatístico da variável dependente e variável independente continua.....	43
Tabela 8: Resumo estatístico das variáveis independentes (dummy) .....	44
Tabela 9: Resumo dos resultados da experiencia por cenário.....	46
Tabela 10: Resumo do modelo, em termo de medida de desempenho .....	47
Tabela 11: informações da rede neuronal .....	47
Tabela 12: Importância das variáveis independentes.....	48
Tabela 13: correlação entre as variáveis .....	54
Tabela 14: coeficiente das variáveis e o nível de significância .....	57
Tabela 15: Resumo dos resultados empíricos .....	58
Tabela 16: Modelo de RNA Vs. MPH.....	62
Tabela 17: Erro relativo e eficiência dos modelos .....	62



## **LISTA DE ANEXOS**

Anexo 1: Está a esquema da rede, assim como o peso de sinápticos e os seus respetivos sinais .....	73
Anexo 2: Estão os pesos de conexões e o parâmetro de estimação. ....	75



## **LISTA DE SIGLAS**

**ADALINA** - *Adaptive Linear Neuron*

**DARPA** - *Defense Advanced Research Projects Agency*

**CVE** – Escudo Cabo-Verdiano

**INE** - Instituto Nacional de Estatística

**IA** - Inteligência Artificial

**IFH** - Imobiliária, Fundiária e Habitat, S.A

**MPH** – Método de Preço Hedónico

**MLP** - *Perceptron* de Múltiplas Camadas

**RNA** – Redes Neurais Artificiais



# 1. INTRODUÇÃO

Em muitos países o mercado imobiliário representa um pilar importante na economia nacional, uma vez que gera emprego e rendas que contribuem significativamente para o aumento dos recursos financeiros em diversos setores.

Kim e Park (2005) definiram o mercado imobiliário como sendo o lugar onde o serviço de habitação é atribuído pelo mecanismo de oferta e da procura. O mercado imobiliário pode ser considerado um local, nacional ou internacional, onde é possível fazer a transação de bens e serviços entre as partes interessadas, através do preço (Pinto, 2013).

Os bens imóveis são considerados um dos gastos domésticos com o preço mais elevado. A alteração dos preços tem sido um dos motivos de preocupação dos indivíduos e do Governo, uma vez que esta alteração influencia as condições socioeconómicas da população (Kim & Park, 2005).

Desde há muito tempo que o setor imobiliário tem contribuído para o crescimento da economia de Cabo Verde, principalmente nas vertentes: imobiliário residencial e imobiliário turístico. O vertente imobiliário turístico engloba empreendimentos de alojamento (casas, apartamentos e hotéis), infraestruturas e serviços de apoio (clínicas), serviços de entretenimento e lazer (campos de golfe, centros hípicas, centros comerciais), essas condições servem de incentivo para a permanência de estrangeiros em território nacional. Cabo Verde é um país que apresenta inúmeras condições que favorecem a prática de investimento na vertente imobiliário turístico nomeadamente, a baixa taxa de imposto sobre o património, o clima ameno e o baixo custo de mão-de-obra.

De acordo com o último censo realizado pelo INE em 2010, em Cabo Verde cerca de 20,53% dos alojamentos familiares existentes não são habitados, o que pode estar relacionado com o elevado número de Cabo-Verdianos a viver fora do país de origem.

O direito à habitação está previsto no artigo 71º, nº 1 da constituição da República de Cabo Verde “ Todos os cidadãos têm direito a habitação condigna”. Para assegurar este direito o Governo estabeleceu uma parceria com a agência imobiliária IFH (Imobiliária, Fundiária e Habitat, S.A) em que elaboraram um plano de ação designado de “Programa Casa para Todos” para promover a habitação social em todo o país, reduzindo assim o défice habitacional. Este programa possibilita que todos os cidadãos tenham acesso a uma habitação sem se preocuparem com os custos associados á aquisição do terreno, construção da moradia, etc. Neste programa o individuo pode proceder ao pagamento do mesmo, de acordo com a sua capacidade financeira.

A Ilha do Sal é a terceira mais populosa do arquipélago. É também uma das ilhas onde os produtos imobiliários residenciais turísticos são mais valorizados em Cabo Verde. Segundo o estudo desenvolvido pela Prime Yield Cabo Verde, a empresa responsável pela avaliação e consultoria imobiliária no âmbito dos Fundos de Investimento Imobiliário em Cabo Verde no ano de 2014, constatou-se que na ilha do Sal o valor médio de vendas de habitações, no que diz respeito ao turismo residencial, é o mais elevado do país (1.882,00 €/m<sup>2</sup> nos apartamentos de tipologia T2 novos e 1.893,00 €/m<sup>2</sup> no caso de moradias de tipologia T3 novas).

No entanto, segundo a Prime Yield Cabo Verde verificou-se uma diminuição do valor médio de vendas de apartamentos T2 novos no ano de 2014 (vendas igual a 1.882,00 €/m<sup>2</sup>) comparativamente ao ano de 2013 (vendas igual a 1.941,00 €/m<sup>2</sup>). Registou-se, ainda, um decréscimo do valor médio de vendas correspondentes há moradias T3 novas, de 2.036,00 €/m<sup>2</sup> em 2013 para 1.893,00 €/m<sup>2</sup> em 2014.

Esta dissertação centra-se na análise dos fatores que influencia o preço das habitações, na Ilha do Sal, em Cabo Verde. Com este estudo espera-se cooperar positivamente na tomada de decisão dos indivíduos e agentes imobiliários no que diz respeito à compra e venda dos imóveis.

Para o melhor desempenho do modelo o número da amostra deve ser significativo. Neste sentido, julgamos que a amostra utilizada (1092 habitações) é uma boa amostra, embora o modelo apresente poucas variáveis explicativas, comparando com os estudos de Carreira (2011), Lam, Yu e Lam (2008) e Teixeira (2011). Com base nesta amostra, pretende-se desenvolver um modelo de redes neuronais artificiais para o mercado imobiliário na ilha

do Sal, em Cabo Verde, com o objetivo de determinar quais as características que mais influenciam o preço de habitações. Foram consideradas características tais como: área, número de quartos, existência de varandas e existência de terraço, número de casas de banhos, localização do imóvel, número de andares e a proximidade com instituições públicas. Para além da análise baseada no desenvolvimento do modelo de redes neurais, efetuou-se a análise pela estimação do modelo de preço hedónico, de forma a realizar uma comparação entre os dois modelos.

A presente dissertação divide-se em cinco secções. Na primeira secção faz-se uma pequena introdução; na segunda secção apresenta-se a revisão da literatura, concretamente as características do mercado imobiliário, uma breve revisão dos métodos de avaliação imobiliário e por fim os contributos empíricos da utilização de redes neurais artificiais (RNA) no sector imobiliário; na terceira secção descreve-se a metodologia, nomeadamente a amostra e variáveis utilizadas, a justificação da utilização do modelo, a forma funcional utilizado no MPH, processo de estimação do preço das habitações pela rede, e por fim o *software* utilizado e as medidas comumente utilizadas para analisar o desempenho da rede; na quarta secção promove-se a apresentação e discussão dos resultados, concretamente as estatísticas descritivas e os principais resultados obtidos para os dois modelos utilizados, e uma comparação de desempenho dos modelos; por último apresentam-se as principais conclusões do estudo, as limitações encontradas e sugestões para investigação futura.



## **2. REVISÃO DE LITERATURA**

Neste capítulo apresenta-se uma revisão ao estado da arte no que diz respeito aos métodos da avaliação do mercado imobiliário, particularmente no que diz respeito à aplicação do modelo de redes neuronais artificiais no mercado imobiliário. Neste sentido, define-se o mercado imobiliário e apresentam-se as suas características. Adicionalmente, efetua-se uma breve revisão dos métodos de avaliação imobiliário, em especial a análise do modelo de preços hedónicos e uma perspetiva histórica do modelo de redes neuronais artificiais. Finalmente, apresenta-se a evidência empírica da aplicação do modelo de redes neuronais artificiais no mercado imobiliário.

### **2.1. Características do mercado imobiliário**

O mercado imobiliário apresenta um conjunto de características que o diferenciam de outros mercados de bens e serviços, o que o torna único. As características deste mercado são assim definidas com base no estudo de Neves, Montezuma e Laia (2010):

- Elevada durabilidade da construção de habitação que leva a que a substituição deste tipo de bens seja mínima;
- Fixidez espacial - esta característica explica a importância da localização na definição do valor da habitação. Os fatores exógenos dos bens, como por exemplo o acesso às infraestruturas e à qualidade da zona, influenciam significativamente o valor de uma habitação;
- Grande heterogeneidade - não existem habitações iguais, existe sempre algum fator de diferenciação, como por exemplo a localização, a qualidade de construção, a idade e o nível de manutenção. A heterogeneidade destes bens contribui para inelasticidade da oferta;
- Falta de liquidez - o vendedor e o comprador antes de realizarem a transação imobiliária necessitam de despender tempo e dinheiro. Existem vários fatores que contribuem para falta de liquidez como:
  - Custos legais e burocráticos;
  - Fixidez;

- Heterogeneidade dos bens imobiliários.

O mercado imobiliário tem outras características que o diferencia dos outros ativos financeiros, tais como, elevados custos de informação e transação, elevado preço de cada unidade, e a intervenção governamental.

## **2.2. Breve revisão dos métodos de avaliação imobiliário**

Pagourtzi, Assimakopoulos, Hatzichristos e French (2003) distinguiram dois tipos de métodos de avaliação do valor do mercado imobiliário: os métodos tradicionais e os métodos avançados. Para estes autores os métodos tradicionais consistem na comparação direta, ou então a partir da determinação de valores através de um modelo de regressão. Por outro lado, os métodos avançados tentam simular de forma indireta o comportamento dos participantes no mercado, a fim de estimar o preço de transação. Estes modelos tendem a ser mais quantitativos.

### **2.2.1. Métodos de avaliação tradicionais**

- **Método comparativo** - é o método de avaliação mais utilizado. A utilização deste método pressupõe a existência de um mercado imobiliário ativo.

Para Pagourtzi *et al.* (2003) o valor de habitação a ser avaliado é determinado a partir da análise do preço de vendas das habitações semelhantes. Na aplicação desse método as vezes é necessário que o avaliador ajuste o preço de vendas para comparar imóveis cujas características sejam diferentes, nomeadamente em relação à idade, qualidade da construção, tamanho, área, localização geográfica, data de transação, vizinhança etc.

A comparação de vendas é fortemente dependente da disponibilidade, exatidão, integridade e atualidade dos dados de transação.

Para eles o procedimento de análise de vendas comparáveis pode ser visto como um processo de quatro etapas: 1) análise de informações sobre operações recentes de imóveis,

semelhantes e comparáveis; 2) ajustamento do preço de vendas dos valores comparáveis para coincidir com as características das habitações; 3) estimação do valor do mercado; 4) apresentação de resultados de forma clara e de fácil compreensão.

Segundo Couto, Manso e Soeiro (2006) a utilização deste método apresenta vários riscos tais como a interferência de critérios subjetivos na avaliação, que é responsável pelas diferenças dos valores estimados pelos avaliadores e à insuficiência de transações de bens semelhantes. Estes riscos conduzem a uma caracterização inadequada do comportamento do mercado e uma avaliação fundamentada de forma insatisfatória. Este método deve ser aplicado apenas quando a informação disponível for suficiente e credível, sendo mais adequado para avaliação de imóveis transacionados com mais frequência como habitações e terrenos e menos adequados no caso de armazéns e unidades industriais.

- **Método de rendimento** - é também conhecido como o método de capitalização de exploração analítico ou indireto. Este método baseia-se no conceito de valorização do imóvel pela capacidade que este possui para gerar rendimentos.

De acordo com Pagourtzi *et al.* (2003) o rendimento representa o retorno ou juro sobre o investimento do imóvel, ou seja o proprietário pode ocupar o imóvel ou simplesmente pode optar por passar o direito de ocupação a um terceiro, onde o inquilino irá pagar uma renda ao proprietário.

Segundo Couto *et al.* (2006) o método de rendimento é essencialmente apropriado para a estimação de valores de propriedades que podem fornecer um rendimento periódico (mensal, sazonal ou anual). Este método depende do rigor na estimação dos rendimentos e do grau de adequação da taxa fixada

- **Método de lucro** - é o rendimento gerado pelo imóvel, onde a propriedade é analisada como uma unidade de produção pelos avaliadores tendo em conta as receitas esperadas do imóvel, não incluindo os custos direto como, o custo com serviços, lavandarias entre outros, para a aquisição desses rendimentos (Pagourtzi *et al.*, 2003).
- **Método residual** - é considerado por Pagourtzi *et al.* (2003) o melhor método para a determinação do valor de um terreno. A estimação deste valor só deve ser feita caso o terreno se encontrar disponível para o seu uso. O cálculo do valor do

terreno é efetuado com base no critério residual, que consiste na dedução dos custos (demolição do edifício existente, custos de construção, honorários profissionais, custos de financiamento, etc.) no valor total do imóvel.

- **Método de custos** - consiste na estimação do custo da reconstrução de uma habitação com as mesmas capacidades, utilidades ou rendimentos do imóvel antigo.
- **Método de regressão múltipla** - pretende relacionar o valor de uma habitação com as diversas características do mesmo, isto é, o preço da habitação é determinado com base nas suas características.
- **Método de regressão *Stepwise*** - este modelo irá incluir todas as variáveis, eliminando sucessivamente o menos importante, ou poderá adicionar as variáveis independentes que apresentam maior coeficiente de correlação com a variável dependente, tendo em conta as variáveis já existentes no modelo.

## 2.2.2. Métodos de avaliação avançados

### 2.2.2.1. Modelo de preço hedónico

O modelo de preço hedónico (MPH) analisa a variação dos preços da habitação em função das suas características.

Segundo Bao, Wang e Li-heng (2010), o MPH foi proposto pelo Rider em 1967 que constatou que a qualidade de algumas características ambientais como o ar, água entre outros, afetam o preço do terreno enquanto bens de consumo e fatores de entrada de produção.

O MPH é composto por duas funções: a função utilidade e a função preço ( Bao *et al.*, 2010).

$$Z = \{z_1, z_2 \dots z_n\} \quad (1)$$

$$P(Z) = p \{ z_1, z_2 \dots z_n \} \quad (2)$$

Onde:

Z – é o vetor das características (estruturais, regionais, de fator individual) da habitação

P (Z) – é o preço correspondente a cada característica

A literatura aponta para a existência de alguns estudos utilizando o modelo do preço hedônico na estimação do preço do imóvel. A sua aplicação é especialmente importante pois permite que se estime o valor de uma variável (dependente) com base num conjunto de outras variáveis (independentes). Quanto mais significativo for o peso de uma variável isolada, ou de um conjunto de variáveis explicativas, tanto mais se poderá afirmar que alguns fatores afetam mais o comportamento de uma variável de resposta do que os outros.

Para estimar o preço do imóvel a partir do MPH, é usada a equação de regressão linear, esta equação determina a relação linear existente entre ambas as variáveis. Na equação 3 consta a representação da equação de regressão linear formulada pelo Kamenta em 1988 (Neto, 2006).

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i \quad (3)$$

Onde:

$y_i$ - Variável dependente;

$x_i$ - Variável independente;

$\beta$  - Coeficiente que determina a explicação da variável dependente pela variável independente;

$\varepsilon_i$ - variável que inclui todos os fatores residuais mais os possíveis erro de medição;

Kong, Yin e Nakagoshi (2007), aplicaram o modelo de preço hedônico na avaliação do valor de amenidade do espaço verde na cidade de Jinan, na China, em que analisaram 124

habitações. Os resultados confirmaram o impacto positivo de amenidade de proximidade aos espaços verdes sobre o preço do imóvel e destacaram uma preferência dos proprietários de imóvel na cidade de Jinan. Os resultados são também importantes para os decisores políticos envolvidos no planeamento urbano.

Tarré (2009), utilizou o modelo hedónico em duas zonas distintas da cidade de Lisboa onde denominou de zona A – Freguesias de Benfica e São Domingos de Benfica e zona B – Freguesias da Lapa, Santo Condestável e Santa Isabel. Concluiu que em cada zona existem variáveis com capacidade explicativa diferente, nomeadamente a variável novo/usado, na zona A, e a área da arrecadação, na zona B. O autor também verificou que algumas variáveis explicativas que são comuns a ambas das zonas, tal como o número de estacionamento e a especificidade do imóvel (esta variável consiste em diferenciar, elementos da amostra que sejam imóveis em condomínios fechados ou que estejam inseridos em edifícios com elevada componente histórica ou por fim que tenham uma localização incomparável).

No estudo de Selim (2008), foi utilizado o modelo hedónico para determinar o preço de renda das casas na Turquia. Foram analisados os dados relativos ao orçamento familiar, realizado em 8600 dos agregados da amostra, no período de 1 de janeiro de 2004 a 31 dezembro de 2004. Os resultados revelaram que entre as variáveis independentes analisadas, a existência de piscina, o tipo de habitação, o número de quartos, a dimensão da habitação, a localização, as características e tipos do edifícios são as variáveis mais significativas que afetam os preços das habitações.

Carreira (2011) desenvolveu o MPH para o mercado imobiliário do distrito de Leiria, de forma a verificar as características de uma habitação que mais influenciam o preço. A amostra foi constituída por 4022 casas com os preços previstos de venda divulgados no portal casa sapo, entre o período de outubro de 2010 a dezembro de 2010. Os resultados indicaram que o preço de uma habitação é fortemente influenciado por algumas variáveis de localização e também por variáveis estruturais. Algumas dessas variáveis têm um efeito positivo sobre o preço da habitação, tais como a localização de moradia no concelho de Óbidos e o número de quartos, especificamente os que possuem quatro ou cinco quartos. Outras variáveis têm um efeito negativo no preço, tais como o estado do uso da habitação, ou seja, nova/usada e a localização da habitação nos concelhos de Marinha Grande e de Leiria.

A Tabela 1 mostra-nos as hipóteses estabelecidas para a influência de algumas características da habitação sobre o preço, assim como as evidências empíricas correspondentes.

Tabela 1: Resumo da evidência empírica de alguns estudos que utilizam o MPH

Hipótese Estabelecida	Evidência empírica
<b>Caraterísticas estruturais</b>	
<b>Área</b>	<b>Sim:</b> Carreira (2011); Teixeira (2011)
	<b>Sim:</b> Carreira (2011); Peddy (2011)
<b>Preço aumenta quando aumenta a área</b>	<b>Não:</b> Limsombunchai, Gan e Lee (2004)
<b>Números de quarto</b>	
<b>Preço aumenta quando aumenta o número de quartos</b>	<b>Sim:</b> Carreira (2011) <b>Nenhuma evidência:</b> Limsombunchai <i>et al.</i> (2004)
<b>Números de casas de banhos</b>	
<b>Preço aumenta quando aumenta o número de casas de banhos</b>	<b>Sim:</b> Limsombunchai <i>et al.</i> (2004) <b>Nenhuma evidência:</b> Carreira (2011)
<b>Existência de varandas e terraço</b>	
<b>Preço aumenta se existir terraço</b>	<b>Sim:</b> Carreira (2011); Mccluskey <i>et al.</i> (2013)
<b>Caraterísticas de localização</b>	
<b>Localização da habitação</b>	
<b>Aumenta o preço dependendo da localização da habitação</b>	<b>Sim:</b> Carreira (2011); Peddy (2011); Selim (2009) <b>Não:</b> Limsombunchai <i>et al.</i> (2004)
<b>Características de vizinhança</b>	
<b>Perto instituição pública</b>	
<b>Aumenta o preço quando a habitação está localizada perto instituição pública</b>	<b>Sim:</b> Carreira (2011) <b>Não:</b> Limsombunchai <i>et al.</i> (2004)

**Nota:** Os estudos que fornecem evidência significativa da hipótese estabelecida aparecem após a palavra Sim, por outro lado, os que fornecem evidência significativa, mas são contrários à previsão teórica aparecerem após a palavra Não, e os estudos que não suportam a previsão teórica aparecem após a palavra Nenhuma evidência.

### 2.2.2.2. Modelo de lógica vaga (Fuzzy Logic) e Modelos de análise espacial

O modelo de lógica vaga é o modelo que se baseia no conhecimento ou nas opiniões.

O principal objetivo do modelo de lógica de *fuzzy* é a transformação de uma informação qualitativa em quantitativa.

Diferente da lógica *booleana* clássica que admite apenas dois valores o verdadeiro e o falso, a lógica *fuzzy* trata de valores que variam entre 0 (absolutamente falso) e 1 (absolutamente verdadeiro). Na lógica *fuzzy* podem existir valores entre verdadeiro e falso, como por exemplo: (0, 0.5, 1) em que esses valores representam o falso, a meio verdade ou talvez verdadeiro e verdadeiro.

Outro ponto crítico na abordagem do modelo de lógicas *fuzzy* é a definição das regras, estas regras baseiam-se na fórmula “se...então...e”, em que o conhecimento num problema/solução de uma determinada área pode ser representado por um conjunto de regras, como por exemplo:

Se... (problema) ...então... (solução);

Se... (problema1) ...e ... (problema2) ...então... (solução).

**O modelo de análise espacial** é um modelo que utiliza instrumentos de análise espacial. Este método de avaliação é o único que depende da variável tempo. O modelo de análise espacial é essencialmente uma abordagem económica com base em dados de séries temporais (Pagourtzi *et al.*, 2003).

### 2.3. Redes neuronais artificiais (RNA)

Modelo de redes neuronais- o sistema de redes neuronais é um modelo de inteligência artificial que replica o processo de aprendizagem do cérebro humano (Pagourtzi *et al.*, 2003).

Uma rede neuronal é o arquétipo utilizado por biólogos e afins na tentativa de simular e modelar os processos cognitivos da mente. Para distinguir entre o modelo natural (cérebro) e artificial (rede), é utilizado o termo redes neuronais artificiais (RNA) (Cortez, 1997).

Uma RNA implementa um procedimento que relaciona de algum modo informação a entrada (desejada) e informação a saída. Ela executa diferentes tipos de tarefas, nomeadamente: a classificação, ou seja, determinação de categoria de um dado padrão de entrada; a associação que atua como uma memória endereçável, que associa um dado valor de saída a um dado valor de entrada; e por fim a codificação que serve para condensar uma entrada, pela produção de uma saída com uma dimensão reduzida (Cortez, 1997).

O modelo da RNA constitui uma importante técnica estatística não-linear capaz de resolver uma gama de problemas de grande complexidade. A RNA é hoje em dia utilizada em diferentes áreas de conhecimento, nomeadamente, filosofia, estatística, psicologia, engenharia, biologia entre outros.

### **2.3.1. Perspetiva histórica do modelo das redes neuronais artificiais segundo (Fausett, 1994)**

O grande interesse no uso da RNA iniciou-se em 1940 com os estudos de McCulloch e Pitts em 1943, reconheceram que a combinação de muitos neurónios simples em sistemas neuronais constituem uma fonte de maior poder computacional. Eles definiram uma função lógica simples e particular com diferentes neurónios em diferentes funções, esta lógica vincula o estudo neurobiológico e matemático (Teixeira, 2011).

Em 1949 o psicólogo Donald Hebb projetou a primeira lei de aprendizagem de RNA. A sua premissa era que, se dois neurónios estivessem ativos simultaneamente, a força da conexão entre eles deve ser aumentado.

Marvin Minsky, em 1951 construiu o SNARC, o primeiro simulador de cadeia neural. Já em 1956, Minsky, juntamente com o McCarthy, Rochester e Shannon, organizaram a primeira conferência sobre Inteligência Artificial (IA), onde foi apresentado um modelo de rede neural artificial. Esse modelo consistia numa simulação de centenas de neurónios

interconectados através de um sistema que verificaria como a rede responderia aos estímulos ambientais (Teixeira, 2011).

Em 1959, Rosenblatt, baseou-se nas ideias de Hebb, McCulloch e Pitts, em que descreveu o primeiro modelo de redes neuronais, denominado de *Perceptron*. O *perceptron* é um sistema de arquivamento que pode identificar padrões e formas geométricas abstratas. Muitos consideram Rosenblatt como o fundador da neurocomputação pelas suas pesquisas e inúmeras contribuições técnicas (Fausett, 1994).

Em 1962, Widrow desenvolveu um tipo diferente de processador para redes neurais, o ADALINE (*Adaptive Linear Neuron*). Este processador dispunha de uma poderosa estratégia de aprendizagem supervisionado que assentava na correção de erros, uma classe muito importante das redes neurais artificiais com uma ampla aplicação hoje (Fausett, 1994).

Em 1969, Minsky e Seymour Papert realizaram um estudo onde mostraram que o modelo *Perceptron* apresentado por Rosenblatt não era muito promissor, devido ao uso de técnicas empíricas, das grandes dificuldades da matemática envolvidas e dos poucos recursos computacionais disponíveis na época. Por força desse trabalho os autores evidenciaram as limitações do *Perceptron* e as pesquisas em redes neurais foram abandonadas quase completamente (Fausett, 1994; Teixeira, 2011).

Fausett considerou a década de 70 um dos períodos em que a pesquisa da RNA teve ligeiros progressos.

Em 1974, Werbos lançou as bases do algoritmo *back-propagation* (retro propagação do erro), que permitiu que redes neurais de múltiplas camadas mostrassem capacidade de aprendizagem (Fausett, 1994).

No início da década de 80, muitos pesquisadores publicaram inúmeras propostas para a exploração de desenvolvimento e pesquisa em redes neurais. O físico e biólogo de John Hopfield em 1982 interessou-se pela neurocomputação. Escreveu vários artigos que levaram vários cientistas a unirem-se nesta nova área emergente. Hopfield reascendeu as pesquisas em neurocomputação, criticando fortemente as teorias apresentadas por Minsky e Papert em 1969 (Fausett, 1994).

No período de 1983-1985, Robert e Todd desenvolveram o programa da DARPA (*Defense Advanced Research Projects Agency*) que fomentou pesquisas de neurocomputação da DARPA, mesmo contrariando todos os princípios (Fausett, 1994).

Em 1988, Broomhead e Lowe descreveram as redes de funções de base radicais ao proporcionaram as redes de múltiplas camadas (Teixeira, 2011).

As razões que levaram à retoma de interesse pelas pesquisas de redes neuronais reemergiram não só pela necessidade de resolver o problema de como treinar uma rede de múltiplas camadas, mas também para melhorar a capacidade computacional.

As redes neuronais artificiais apresentam um conjunto de vantagens na sua utilização, 1) a habilidade da RNA de aprender automaticamente o mapeamento desejado entre as entradas e as saídas, através de um processo iterativo de ajustes aplicados aos seus parâmetros livres (ex.: pesos sinápticos); 2) a capacidade que as RNA possuem de adaptar seus pesos sinápticos perante as modificações no meio ambiente; 3) a tolerância a falhas, o conhecimento é distribuído pela rede, desta forma, uma parte das conexões podem estar inoperante, sem mudanças significativas no desempenho de toda a RNA, e 4) a resposta a evidências, na sua utilização como classificadora de padrões, uma RNA pode fornecer, em sua saída, não só a informação relativa a um conjunto de entrada pertencente ao modelo, mas também uma informação sobre a confiança nos resultados. Desta forma, essas informações podem ser utilizadas para rejeitar padrões duvidosos (Passari, 2003; Teixeira, 2011).

### **2.3.2. Características de redes neuronais artificiais**

As RNA constituem um método de resolução de problemas através da simulação do cérebro humano, através da aprendizagem, errando e fazendo descobertas (Teixeira, 2011). As RNA apresentam as seguintes características Teixeira (2011):

- As RNA são constituídas por neurónios que são unidades de processamento, estas unidades são simulação das células do cérebro humano, através da recepção e transmissão de informações. Cada unidade possui ligações com outras unidades onde enviam e recebem sinais, muitos desses neurónios operam ao mesmo tempo.

- Uma RNA deve possuir pelo menos duas camadas (entrada e saída), normalmente são utilizadas as redes que detém mais camadas de conexão, denominadas de camadas intermedias. Essa tipologia é denominada de redes de múltiplas camadas. Como por exemplo: numa situação em que uma rede é constituída por três camadas: a primeira camada representa a entrada onde as unidades recebem os padrões; em seguida a camada intermédia onde se encontra o processamento e a extração das características; e a última camada é a camada de saída onde são localizados e apresentados os resultados finais.
- É uma rede altamente interconectada. Assim, devido á sua conectividade as RNA têm propriedades únicas que permitem depois comparar com o modelo do preços hedónicos (Cortez, 1997).
- As RNA atingem são mais eficientes em problemas de classificação ou associação, onde é necessário um certo grau de tolerância a falhas, em que existe muita informação, aos quais não se podem aplicar outros procedimentos (Cortez, 1997).
- A transparência é uma das características de RNA, uma vez que podem ser vistos como uma *caixa negra*, sendo a saída uma função desconhecida. Em princípio a rede neuronal pode imitar qualquer função computacional (Cortez, 1997).

### 2.3.3. Ativação e função de ativação

A RNA é definida como um conjunto de unidades computacionais, chamadas células (nó), que possuem memória local e um conjunto de **conexões** (ou ligações) que unem essas células, essa é uma das formas mais simples de definir RNA. Estas conexões têm em si associados valores numéricos designados de **pesos**. Temporariamente, cada unidade examina as suas entradas e posteriormente é calculado um valor numérico chamado de **ativação**. Este cálculo processa-se em dois estágios distintos, o primeiro estágio consiste na soma dos produtos de sinais de entradas com o peso. Já o segundo estágio consiste na aplicação de uma função de saída chamada função ativação. Por convenção existe uma

célula especial cuja saída é sempre 1 (um) e que está ligada a todas as células, sendo o peso desta ligação é denominado por bias.

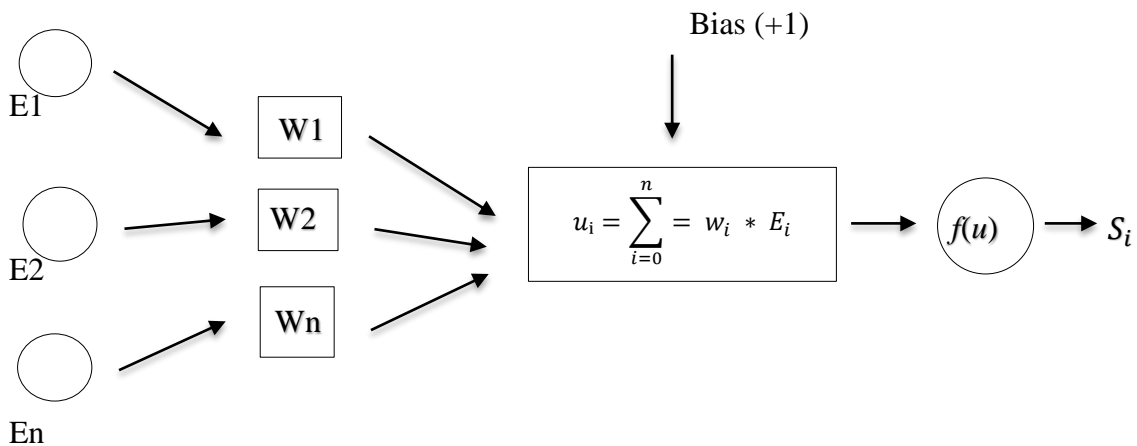


Figura 1. Ativação de neurónio numa RNA

Fonte: Adaptado de Lima, Kimura, Neto e Perera (2010)

Onde:

- (E1, E2...En) - são os valores da entrada;
- (W1, W2...Wn) - são os pesos que cada atributo tem na rede;
- b- bias ou o termo constante da equação;
- $F$  - função ativação;
- $X$  - resultado do somatório entre a entrada e o peso;
- $S_i$  - saída.

Entre as funções de ativação existentes, a função logística é a mais usada, no que se refere a RNA, isto porque, esta função não é limitada. A função logística é uma função crescente, ou seja, o valor da função aumenta com o valor do argumento e é contínua em seu domínio. Na função logística não é necessário o uso do número de Neper no denominador, esse número pode ser substituído por qualquer valor constante maior do que 1, em que quanto maior for o valor da constante, mais rapidamente a função se aproxima aos seus limites (Teixeira, 2011). Esta função é representada da seguinte forma:

$$\frac{1}{1 + e^{(-x)}} \quad (4)$$

Na Tabela 2 estão representados outras funções que podem ser usados nas RNA, estas funções podem ser utilizadas de forma linear e não-linear.

Tabela 2: **Função Ativação da RNA**

Nome	Função F (x)	Contradomínio
<b>Linear</b>	$x$	$] -\infty, +\infty [$
<b>Degrau</b>	$\begin{cases} 1, se x \geq 0 \\ 0, se x < 0 \end{cases}$	$[0, 1]$
<b>Tangente hiperbólica</b>	$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	$[-1, 1]$

Fonte: Adaptado de Teixeira (2011)

#### 2.3.4. Tipologia e treinamento de redes neuronais

Existem várias tipologias de arquitetura de RNA, apresentando cada uma as suas próprias potencialidades. A grande distinção é feita entre as redes recorrentes e a *feedforward*. Na primeira, as conexões podem ser feitas entre quaisquer células, enquanto que numa rede *feedforward* as conexões são unidirecionais, não existindo realimentação, sendo assim formado por camadas, isto é, os resultados de uma célula não influenciam os resultados da outra célula na mesma camada. A rede *feedforward* é a tipologia mais utilizada (Cortez, 1997).

Para adquirir um certo nível de aprendizagem uma rede neuronal tem que submeter-se a um treinamento. O principal objetivo é fazer com que a aplicação de um conjunto de entrada produza um conjunto de saída desejado, ou no mínimo consistente. Uma vez treinada uma rede neuronal, esta deve ser capaz de associar um conjunto de valores para uma saída específica.

Os processos de treinamento geralmente podem ser classificados em dos tipos: supervisionados e não-supervisionados. Para Almeida (1993 como citado em Passari, 2003) o treinamento supervisionado requer a presença de valores de entrada e a correspondente saída, a partir do qual é calculado o erro, que é utilizado para corrigir o peso das conexões das redes. No treinamento não-supervisionado são apresentados

apenas os vetores de entrada, os valores de saída são construídos automaticamente pela rede. Este tipo de treinamento é capaz de capturar regularidade nos valores de entrada, e não recebe informações adicionais.

Petron (1999 como citado em Passari, 2003) identifica diferentes passos no processo de montagem e treinamento de uma rede, tais como:

- A definição da tipologia da rede:
- A rede sofre um processo de aprendizagem, através de informações de uma certa entrada e saída conhecida;
- As conexões entre as células são modificadas num processo iterativo, até que a rede forneça saída comparável com as informações apresentadas. Por isso diz-se que a rede neuronal adquire conhecimentos através da experiência;
- A rede está pronta para fornecer respostas para o novo problema.

O aprendizado pode continuar indefinidamente. O conhecimento está na força de ligações entre neurónios, ou seja no peso dado a cada conexão.

### **2.3.5. Evidências empírica da utilização das Redes Neurais no setor imobiliário**

Nos anos 90 a previsão com a RNA teve um grande progresso, isto devido ao trabalho de, Borst (1991) considerado um dos pioneiros no uso de RNA. Desde então o modelo tornou-se um atrativo no que se refere aos modelos econométricos tradicionais. A vantagem primordial dessa prática é a capacidade de lidar com relações não lineares ou modelos inicialmente desconhecido. Uma RNA é não-linear se esta for constituída de neurônios artificiais também não-lineares. Esta é uma característica importante, pois a maioria dos sistemas físicos responsáveis pela geração do mapeamento entre os sinais de entrada e saída desejada são não-lineares. Os outros modelos econométricos (MPH) têm capacidade de lidar com uma relação linear. Esta é a maior diferença existente entre o modelo de RNA e MPH. Na literatura deparamo-nos com os sucessos e insucessos no uso do modelo de RNA, provavelmente isso ocorre devido a diferentes variáveis das condições de entrada e diferentes mercado.

Logo depois do trabalho de Borst (1991), Do e Grudnitski (1993) utilizaram o modelo de RNA para analisar o efeito da idade no preço das casas. O estudo foi realizado na Califórnia, precisamente no município de San Diego e foram tidas em conta 242 moradias unifamiliares vendidas durante Janeiro de 1991 a Setembro 1991. Os resultados indicam que a idade influencia negativamente o preço das casas nos primeiros 16-20 anos, e depois passará a influenciar positivamente.

Collins e Evans (1994) aplicaram o modelo de RNA no Reino Unido para verificar qual é o efeito dos ruídos das aeronaves nos valores das propriedades residenciais. Analisaram 3.472 dados relativamente a vendas das habitações no período de doze meses (Abril de 1985 a Março de 1986). Os resultados demonstram que o uso de RNA foi um sucesso na análise da influência dos ruídos das aeronaves no preço da habitação.

Worzala, Lenk e Silva (1995) aplicaram a tecnologia de redes neuronais para avaliação imobiliária e compararam o desempenho das redes neuronais e da regressão múltipla, para estimar o preço de venda dos imóveis residenciais em Fort Collins (Colorado, Estados Unidos da América). Os resultados indicam que, o modelo de rede neuronal superou ligeiramente os modelos de regressão múltipla em alguns casos. No entanto, a diferença entre as duas técnicas de modelagem foi muito pequena. Além disso, os resultados entre os dois pacotes neurais que foram usados não eram consistentes. Em todos os testes realizados somente um dos pacotes de *software* de rede neuronal superou o modelo de regressão múltipla.

A RNA é muitas vezes criticada pelo fato de utilizar um julgamento superficial na determinação dos ajustes necessários dos dados, e também por ser impreciso para os avaliadores (Worzala *et al.*, 1995).

Rossini (1997) estabelece uma relação entre o modelo de redes neuronais e MPH. O autor considera que as redes neurais parecem ser uma ferramenta mais adequada a pequenos conjuntos de dados, enquanto que o MPH é claramente superior para conjuntos de dados maiores. Assim, o MPH é estatisticamente pobre com pequenos conjuntos de dados, um problema não encontrado nas redes neurais. O autor acredita que num futuro próximo, com o aumento de ferramentas de cálculo, RNA tornará uma ferramenta de grande poder na estimação do preço das habitações.

Kershaw e Rossini (1999) realizaram um estudo cujo objetivo é verificar se as redes neurais são uma inclusão útil para as ferramentas usadas na indexação de preço de propriedade. Adicionalmente, os autores Worzala *et al.* (1995), Do e Grudnitski (1993), Rossini (1997), Nguyen e Cripps (2001) concluíram que a RNA é uma técnica econométrica superior, quando comparada com o método de preço hedônico.

No estudo de Nguyen e Cripps (2001), os autores compararam o desempenho preditivo entre MPH e RNA para determinar o valor do imóvel residencial familiar. Obtiveram uma amostra de 3.906 observações. Para estes autores, no que diz respeito à especificação funcional o modelo de RNA torna-se mais complexo, sendo que o tamanho da amostra de treinamento deve ser aumentado para que as RNA tenham maior desempenho do que o MPH.

À semelhança com Nguyen e Cripps (2001), Limsombunchai, Gan e Lee (2004) estabeleceram a comparação do poder preditivo do MPH com as RNA na previsão do preço dos imóveis. Para o efeito utilizaram uma amostra de 200 habitações em Christchurch, Nova Zelândia, e tiveram em consideração as seguintes variáveis: o tamanho da habitação, a idade da habitação, o tipo de habitação, o número de quartos, o número de casas de banho, o número de garagens, as facilidades nos acessos e a localização geográfica. Os resultados empíricos demonstram o potencial das redes neurais artificiais na previsão de preço das habitações.

No mesmo ano Ge e Runeson (2004) utilizaram as RNA na previsão do preço dos imóveis residenciais na cidade de Hong Kong, no período compreendido entre o primeiro trimestre de 1981 e o quarto trimestre de 2002. Neste caso os resultados também mostram que a metodologia de rede neural tem a capacidade de aprender, generalizar e convergir séries cronológicas.

Desde a primeira aplicação de redes neuronais no mercado imobiliário, no início dos anos 90, tem-se demonstrado muito interesse na aplicação da rede neural como técnica alternativa de modelagem na avaliação do valor do mercado imobiliário em várias regiões do mundo. Estes resultados sublinham a procura crescente ao longo do tempo, de novos e melhores algoritmos para o funcionamento de redes neurais. Tradicionalmente, o MPH ou outras abordagens baseadas em regressão eram usados na avaliação da propriedade. Na maioria dos estudos realizados até agora efetua-se a comparação entre as duas técnicas

(MPH e RNA). No entanto, a técnica de redes neurais também pode ser vista como complementar às abordagens de MPH ou também como substituta (Kauko, 2003).

Teixeira (2011) identificou duas vantagens principais de sistema de inteligência artificiais, em relação aos métodos tradicionais: (1) os sistemas de inteligência artificial têm, em testes, a média de erro entre 5 e 10%, enquanto os modelos de regressão múltipla têm erros entre 10 e 15%. No entanto, em algumas experiências foram obtidos resultados semelhantes quando as amostras são homogêneas; (2) a capacidade para estimar o valor das habitações que têm características significativamente diferentes. No entanto, em alguns estudos, os sistemas de inteligência artificial têm dificuldade de estimar com precisão os valores das habitações com características especiais.

Lai (2006) aplicou a tecnologia de rede neuronal recorrente “*backpropagation*” para avaliação imobiliária no mercado da cidade de Kaohsiung. No seu estudo empírico o autor reconheceu que um modelo de RNA gera mais erro de previsão do que um modelo MPH, devido à natureza do modelo de rede neural. Os resultados mostram que o desempenho de RNA é melhor do que o modelo de MPH na cidade de Kaohsiung.

Lam, Yu e Lam (2008) realizaram um estudo cujo objetivo foi explorar um modelo matemático eficaz e eficiente na previsão do preço das habitações, de forma a ajudar os compradores nas suas tomadas de decisão no âmbito do mercado imobiliário em Hong Kong. Os resultados da pesquisa demonstram que a RNA é o melhor modelo para a previsão do preço das habitações, quando se consideram as variáveis de entrada apropriadas e amostras de reduzida dimensão.

Selim (2009) utilizou RNA como uma técnica alternativa na estimação do preço das habitações na Turquia. Foram analisados 2004 dados e os resultados demonstraram que a RNA é a melhor alternativa no que diz respeito à previsão do preço das habitações no mercado considerado. Constatou-se, ainda, que a existência de piscina, tipo de habitação, número de quartos, dimensão da habitação, localização, característica e tipo do edifício são as variáveis que mais influenciam o preço das habitações com o uso do MPH. Comparando o desempenho de previsão entre os modelos de RNA e MPH, este estudo demonstra que o modelo de RNA pode ser uma alternativa melhor na previsão do preço das habitações na Turquia.

No estudo de Igbiosa (2011), utilizou o sistema de inteligência artificial para verificar quais das características do imóvel residencial mais influenciavam o preço na Nigéria, mais precisamente na cidade de Benim e Lagos. Esse estudo revelou que existe uma correlação linear e positiva entre características e o preço do imóvel, também mostra que embora um imóvel apresente várias características, apenas poucos têm uma importância significativa sobre o preço. Assim, algumas características importantes foram identificadas no mercado nigeriano, tais como: tipo de imóvel (moradia unifamiliar ou apartamento), atratividade do bairro (bairro altamente, bastante ou pouco atraente), número de casas de banho, dimensão da habitação, o ano em que o imóvel foi vendido ou avaliado e a categoria de bairro (suburbana, exclusiva, baixa densidade ou favela). Os resultados demonstram claramente que, de facto, as redes neurais artificiais poderiam tornar-se um modelo prático e dinâmico na avaliação justa das unidades de habitação no mercado imobiliário nigeriano. O autor acredita firmemente que o conhecimento das características influenciará significativamente no preço e na qualidade da propriedade.

No mesmo ano Hamzaoui, Alfredo, e Perez (2011) aplicaram a abordagem das RNA para desenvolver um modelo matemático que prevê o preço de venda dos imóveis residenciais. O estudo baseia-se na avaliação das vendas de habitações em Marrocos, precisamente em Casablanca, onde foi formada uma rede *feedforward*, com uma camada oculta. A RNA foi obtido por 148 conjuntos de padrões de entrada e saída, aplicando o algoritmo de *backpropagation*. Para as redes foram aplicados: a função logística, tangente hiperbólica e a função de transferência linear. Os dados de treinamento foram obtidos de uma arquitetura de RNA, composta por cinco neurónios na camada oculta, o que torna possível prever o preço de venda das habitações. Do modelo resultaram boas previsões com alto coeficiente de determinação ( $R^2 = 0.952$ ).

Os autores sugeriram que o novo modelo de RNA poderia ser usado como uma ferramenta de previsão fiável do preço de venda das habitações. O preço de venda dos imóveis residenciais durante o processo de avaliação foi previsto com sucesso, aplicando uma rede neuronal de três camadas, com cinco neurónios na camada oculta e usando *backpropagation* do algoritmo de *Levenberg Marquardt*. O estudo foi realizado a fim de estimar o comportamento do sistema sob diferentes variáveis. Todos os parâmetros estudados neste trabalho (por exemplo, localização, proximidade urbana de referência, número de quartos, número de casas de banho, número de estacionamento, existência de

elevador, data de avaliação, estilo de construção, etc.) têm efeitos consideráveis sobre o preço de venda.

Os resultados da modelagem confirmaram que o modelo de rede neuronal pode reproduzir com eficácia os dados experimentais e procurar as relações entre todas as variáveis, tanto qualitativas como quantitativas, sem usar modelos matemáticos difíceis e rígidos como o algoritmo de regressão, a rede não é obrigada a seguir uma série de instruções. No entanto, a rede funciona com base na aprendizagem, cria as suas próprias “regras” e aprende com seus erros para prever o comportamento do processo com erro mínimo possível e obter a solução ideal do problema.

No trabalho de Chiarazzo, Caggiana, Marinellia e Ottomanellia (2014) avaliou-se o desempenho de RNA na estimação do preço de vendas de imóveis na cidade de Taranto (Itália), esse estudo analisa o impacto das condições ambientais (poluição e paisagem) no mercado imobiliário. Consideram um conjunto de habitação na área urbana de Taranto, onde se situam a maior fábrica de aço Europeu e o segundo porto industrial. Além disso, também realizaram uma análise de sensibilidade para identificar as variáveis de entrada mais significativa. Tanto os modelos de regressão múltipla como os modelos de rede neuronal, foram úteis em destacar como diferentes variáveis de transporte (número de autocarros de linhas internas, servindo a zona e propriedade situando 500 m da estação de comboio suburbano), bem como a qualidade ambiental que afeta os preços dos imóveis. Os resultados fornecidos pela rede neuronal podem apoiar os investimentos no sistema de transportes, e podem ajudar os avaliadores a fazer avaliações e regeneração ambiental.

Os autores Tajani, Morano, Locurcio e Addabbo (2015) utilizaram RNA para determinar o valor das propriedades em tempos de crise. Os autores verificaram a necessidade de utilizar um modelo capaz de operar mesmo com dados limitados, e também capaz de capturar a relação entre as variáveis explicativas e o preço de venda, com o objetivo de calcular o valor do imóvel no curto prazo. Além de usar RNA os autores também utilizaram a técnica de regressão polinomial evolutiva (EPR). Nesse trabalho, os métodos de RNA e o procedimento EPR são usados para a construção de modelos de estimativas de valores das propriedades. Assim, utilizaram uma amostra de 90 apartamentos vendidos num distrito da cidade de Bari (Itália), os dois modelos foram implementados e ambos mostraram um desempenho excelente, mas o modelo EPR-MOGA  $M_3$  permitiu obter

simultaneamente a melhor precisão estatística na previsão dos preços de mercado, uma verificação rápida da consistência empírica dos resultados obtidos, bem como a superação das principais limitações das RNA.

Na verdade, as RNA foram consideradas como “*caixas negras*”, ou seja, não permitem gerar uma relação funcional e direta entre os valores de entrada e de saída, além disso, pode acontecer que os resultados obtidos através de RNA não poderiam ser estáveis, mas poderiam melhorar com o aumento do tamanho da amostra enquanto que o modelo EPR, supera estas deficiências, a transparência da expressão matemática obtida permite verificar a importância das variáveis explicativas (e quantificar) na formação de preços da propriedade (Tajani *et al.*, 2015).

A Tabela 3 mostra o resumo da revisão de literatura com o uso de redes neurais artificiais no mercado imobiliário.

**Tabela 3: Resumo de evidencia empírica que utiliza a metodologia das RNA no setor imobiliário**

<b>Tema do Artigo</b>	<b>Autor/ Ano</b>	<b>País</b>	<b>Número da Amostra</b>	<b>Resultado do estudo</b>
A análise de redes neurais no efeito da idade sobre o preço de habitação	Do e Grudnitski (1993)	Município de Sam Diego (Califórnia)	242	Positivo
O ruído dos aviões e valor de propriedade residenciais- abordagem de redes neurais artificiais	Collins & Evans (1994)	Reino Unido	3.472	Positivo
Uma exploração de redes neurais e á sua aplicação na avaliação de imóveis	Worzala <i>et al.</i> (1995)	fort collons	288	Negativo
Aplicação de redes neurais artificiais para a valorização do imóvel residencial	Rossini (1997)	Austrália	223	Positivo
Uso de redes neurais para estimar os índices de preços de qualidade constante das habitações.	Kershaw e Rossini (1999)	Austrália (Port Pirie, Stirling e Unley)	_____	Positivo

Tabela 3: **Resumo de evidencia empírica que utiliza a metodologia das RNA no setor imobiliário**  
(cont.)

<b>Tema do Artigo</b>	<b>Autor/ Ano</b>	<b>País</b>	<b>Número da Amostra</b>	<b>Resultado do estudo</b>
Previsão do valor de habitação: Uma comparação de análises de regressão múltipla e redes neurais artificiais	Nguyen e Cripps (2001)	Tennessee (USA)	3.906	Positivo
Aplicações de redes neurais atuais, que envolvem a modelagem espacial de preços dos imóveis	Kauko (2003)	_____	_____	Positivo
Previsão de preço das casas: Modelo preço hedônico Vs. Rede Neural Artificial	Limsombunchai <i>et al.</i> (2004)	Christchurch (Nova Zelândia)	200	Positivo
Modelagem de preços de propriedade através Modelo de Rede Neural em Hong Kong	Ge e Runeson (2004)	Hong Kong	_____	Positivo
Análise do modelo de avaliação em massa utilizando redes neurais artificiais na cidade de Kaohsiung	Lai (2006)	Kaohsiung	2.471	Positivo
Determinante de preços das habitações na Turquia: regressão hedônica Vs. Rede Neural Artificial	Selim (2009)	Turquia	2004	Positivo
Determinantes do valor do imóvel residencial na Nigéria – uma abordagem de Rede Neural	Igbinsosa (2011)	Nigéria	3034	Positivo
Aplicação de redes neurais artificiais para prever o preço de venda no processo de avaliação imobiliária	Hamzaoui <i>et al.</i> (2011)	Marrocos	148	Positivo

Tabela 3: **Resumo de evidencia empírica que utiliza a metodologia das RNA no setor imobiliário** (cont.)

<b>Tema do Artigo</b>	<b>Autor/ Ano</b>	<b>País</b>	<b>Número da Amostra</b>	<b>Resultado do estudo</b>
Um modelo baseado em rede neural na estimativa de preços de imóveis, considerando a qualidade ambiental e a localização da propriedade	Chiarazzoa <i>et al.</i> (2014)	Itália	193	Positivo
Avaliação de imóveis em tempos de crise. Redes neurais artificiais Algoritmos evolutivos em comparação	Tajani <i>et al.</i> (2015)	Itália	90	Positivo

**Nota.** Os resultados dos estudos são identificados como “positivo” quando o uso das RNA atinge um bom desempenho, ou seja, o resultado esperado foi atingido; quando os resultados do estudo é identificado como “negativo” significa que os resultados apresentam um mau desempenho, ou seja, os resultados esperados não foram atingidos.

Nos estudos expostos têm sido usados diferentes amostras e variáveis. Na maioria dos estudos as RNA têm sido consideradas uma boa alternativa na estimação do preço das habitações, assim com o MPH e outras abordagens baseadas na regressão.



### 3. METODOLOGIA

Neste capítulo procede-se ao desenvolvimento do estudo empírico no âmbito dos objetivos deste trabalho. Numa primeira parte será apresentada a amostra; na segunda parte procede-se à identificação e descrição das variáveis usadas no presente estudo com o modelo de RNA e MPH; na terceira parte a forma funcional utilizada no MPH e a equação do modelo; na quarta parte a descrição do processo de estimação do preço das habitações pela rede neural; na quinta parte é apresentada a justificação da utilização do modelo de RNA, e finalmente, efetua-se a identificação do *Software* utilizado e das medidas de desempenho comumente utilizadas para aferir do desempenho de uma rede.

#### 3.1. Apresentação da Amostra

Os dados utilizados neste estudo foram fornecidas por uma agência imobiliária em Cabo Verde. A amostra inicial era constituída por 9534 registos. Todavia, nestes registos não estavam representadas todas as variáveis necessárias para este estudo, uma vez que a maioria dos dados representava a venda de espaços comerciais e terrenos o que não é o objetivo do estudo. Assim, selecionando apenas os registos com todas as variáveis pertinentes, contando agora com 1200 observações, incluindo os *outliers*.

Os *outliers* identificados são os que dizem respeito a variável “preço de venda” e a “área”, quando analisados isoladamente. Foram considerados *outliers* as observações cujo valor estivesse fora do intervalo:

[1º quartil – 1,5 \* amplitude interquartis; 3º quartil + 1,5 \* amplitude interquartis]

Com o resultado foram identificados 108 *outliers*, que posteriormente foram eliminados. Assim, a amostra final do estudo é constituída por 1092 habitações vendidas na Ilha do Sal, entre 2009 e 2014. A informação fornecida pela imobiliária permite-nos usar o valor real do preço das vendas. No que diz respeito às características de vizinhança, os dados foram recolhidos no terreno, tendo como referência o nome das cidades e das zonas da Ilha do Sal.

A Tabela 4 mostra as cidades e as zonas da Ilha do Sal ordenados alfabeticamente e as respetivas distribuições da amostra.

Tabela 4: Distribuição da amostra pelas cidades e zonas da Ilha do Sal

Cidades/Zonas	Número de observação
Algodoeiro- Santa Maria	4
Bairro novo- espargos	10
Bairro Novo I- Espargos	2
Bairro Novo II- Espargos	9
Cidade de Santa Maria	906
Chã de Matias- Espargos	4
Entrada de Santa Maria	24
Hortelão- Espargos	8
Leste de Santa Maria	4
Morro curral- espargos	2
Murdeira	28
Norte de Santa Maria	16
Palmeira	8
Ponta Preta- Santa Maria	2
Praia Antónia Sousa- Santa Maria	45
Preguiça- Espargos	4
Ribeira Funda	5
Tanquinho- Santa Maria	11

A cidade de Santa Maria é a mais representativa da amostra, isto pode estar relacionada com o maior número de concentração da população na cidade, enquanto que a zona de Morro Curral e o Bairro Novo são raramente representados pela amostra.

## 3.2. Identificação e Descrição das Variáveis

### 3.2.1. Variável dependente

A semelhança do estudo de Teixeira (2011), Worzala *et al.* (1995) e Do e Grudnitski (1993) a variável dependente é o preço real da venda das habitações.

### 3.2.2. Variáveis independentes

As variáveis independentes foram escolhidas de acordo com as características das habitações disponibilizadas pela agência imobiliária. Para inclusão no modelo de redes neurais algumas variáveis qualitativas com a existência de varandas, existência de terraços, proximidade com instituições públicas e zona<sup>1</sup> foram introduzidas na rede como variáveis *dummy*, as variáveis como número de quartos, número de casas de banho, localização do imóvel, número de andares e área não sofreram transformações.

Na Tabela 5, encontra-se as variáveis utilizadas no modelo de redes neurais e as descrições dos atributos das habitações e a sua classificação em características estruturais, características de localização e características de vizinhança

Tabela 5: Descrição e classificação dos atributos das habitações (RNA)

Variáveis	Nome das variáveis	Definição	Fonte	Descrição das variáveis dummy	Efeito esperado no preço de venda
<b>Caraterísticas Estruturais</b>					
<b>Área</b>	Tamanho da habitação	Área do imóvel em m <sup>2</sup>	Worzala <i>et al.</i> (1995)	_____	Positivo (+)
<b>N_Q</b>	Número de quartos	Número de quartos na habitação	Do e Grudnitski (1993); Limsombunchai <i>et al.</i> (2004)	_____	Positivo (+)
<b>N_WC</b>	Número de casas de banhos	Número de casa de banhos	Do e Grudnitski (1993); Limsombunchai <i>et al.</i> (2004)	_____	Positivo (+)
<b>TRR</b>	Terraço	Indica se a habitação possui terraço	Collins e Evan (1994)	Dummy= 1 se habitação despõe de um terraço; 0 caso contrario	Positivo (+)

<sup>1</sup> Esta variável foi criada na tentativa de melhorar o desempenho da rede.

Tabela 5: Descrição e classificação dos atributos das habitações (RNA) (cont.)

Variáveis	Nome das variáveis	Definição	Fonte	Descrição das variáveis dummy	Efeito esperado no preço de venda
<b>Caraterísticas Estruturais</b>					
<b>VRR</b>	Varanda	Indica se a habitação possui varanda	Lam <i>et al.</i> (2008)	Dummy= 1 se habitação dispõe de uma varanda; 0 caso contrario	Positivo (+)
<b>ANDAR</b>	Andar	É o andar onde se encontra o imóvel	Lam et al. (2008)	_____	Positivo (+)
<b>Caraterísticas de localização</b>					
<b>Localização</b>	Loc	Localização da habitação na Ilha do Sal <sup>a</sup>	Amaro (2012)	_____	Positivo (+) Negativo (-)
<b>Zona</b>	Zona		Amaro (2012)	Dummy= 1 se a habitação se encontra localizada na cidade de santa maria; 0 caso contrario <sup>b</sup>	Positivo (+) Negativo (-)
<b>Caraterística de vizinhança</b>					
<b>PIP (Finanças, E_S, C_M, HOSP)</b>	Perto instalação pública	Indica-nos se a habitação se encontra perto de escola, câmara municipal, hospital...	Limsombunchai <i>et al.</i> (2004)	Dummy= 1 se a habitação se encontra perto de instalações pública; 0 caso contrario	Positivo (+)

Tabela 5: Descrição e classificação dos atributos das habitações (RNA) (cont.)

**Nota.**<sup>a</sup>Neste caso, usamos variáveis *dummy*, na determinação da localização de habitação onde: S\_M representa Santa Maria; AG\_SM representa Algodoeiro- Santa Maria; BN I\_ES representa Bairro Novo I- Espargos; BN II\_ES representa Bairro Novo II- Espargos; BN\_ES representa Bairro novo-Espargos; CM\_ES representa Chã de Matias-Espargos; ESM representa Entrada de Santa Maria; H\_ES representa Hortelão- Espargos; LSM representa Leste de Santa Maria; MC\_ES representa Morro curral- espargos; NSM representa Norte de Santa Maria; PL representa Palmeira; PP\_SM representa Ponta Preta- Santa Maria; PAS\_SM representa Praia Antónia Sousa- Santa Maria; PG\_ES representa Preguiça- Espargos; RF\_ES representa Ribeira Funda- Espargos; MD representa Murdeira. <sup>b</sup> As zonas são divididas em: Zona\_1 que inclui todas as habitações localizada em Santa Maria (Santa Maria, Algodoeiro-Santa Maria, Entrada de Santa Maria, Leste de Santa Maria, Ponta Preta- Santa Maria, Praia Antónia Sousa- Santa Maria, Tanquinho- Santa Maria); e Zona\_2 que inclui as restantes habitações localizada em qualquer parte da ilha do Sal (Bairro Novo I- Espargos, Bairro Novo II- Espargos, Bairro novo- espargos, Chã de Matias- Espargos, Hortelão-Espargos, Morro curral- espargos, Palmeira, Preguiça- Espargos, Ribeira Funda, Murdeira).

Para inclusão no modelo de preços hedónicos, foram feitas as transformações das variáveis qualitativas em variáveis *dummy*, como por exemplo: número de quartos, existência de varandas, existência de terraços, número de casas de banho, localização do imóvel, número de andares e proximidade com instituições públicas.

Na Tabela 6, encontra-se as descrições dos atributos das habitações e a sua classificação em características estruturais, características de localização e características de vizinhança.

Tabela 6: Descrição e classificação dos atributos das habitações (MPH)

Variáveis	Nome das variáveis	Definição	Fonte	Descrição das variáveis <i>dummy</i>	Efeito esperado no preço de venda
<b>Caraterísticas Estruturais</b>					
Área	Tamanho da habitação	Área do imóvel em m <sup>2</sup>	Worzala <i>et al.</i> (1995)	_____	Positivo (+)
N_Q	Número de quartos	Número de quartos na habitação	Do e Grudnitski (1993); Limsombunchai <i>et al.</i> (2004)	Inclui 3 variáveis <i>dummy</i> , cada uma identificando o número de quartos <sup>a</sup>	Positivo (+)

Tabela 6: Descrição e classificação dos atributos das habitações (MPH) (cont.)

Variáveis	Nome das variáveis	Definição	Fonte	Descrição das variáveis <i>dummy</i>	Efeito esperado no preço de venda
<b>Caraterísticas Estruturais</b>					
<b>N_WC</b>	Número de casas de banhos	Número de casa de banhos	Do e Grudnitski (1993); Limsombunchai et al. (2004)	Inclui 2 variáveis <i>dummy</i> , cada uma identificando o número de casas de banho <sup>b</sup>	Positivo (+)
<b>TRR</b>	Terraço	Indica se a habitação possui terraço ou varanda	Collins e Evan (1994)	<i>Dummy</i> = 1 se habitação dispõe de um terraço; 0 caso contrario	Positivo (+)
<b>VRR</b>	Varanda	Indica se a habitação possui terraço ou varanda	Lam <i>et al.</i> (2008)	<i>Dummy</i> = 1 se habitação dispõe de uma varanda; 0 caso contrario	Positivo (+)
<b>ANDAR</b>	Andar	É o andar onde se encontra o imóvel	Lam <i>et al.</i> (2008)	Inclui 3 variáveis <i>dummy</i> identificando o andar <sup>c</sup>	Positivo (+)

Tabela 6: Descrição e classificação dos atributos das habitações (MPH) (cont.)

Variáveis	Nome das variáveis	Definição	Fonte	Descrição das variáveis <i>dummy</i>	Efeito esperado no preço de venda
<b>Caraterísticas de localização</b>					
<b>Localização</b>	Loc	Localização da habitação na Ilha do Sal	Amaro (2012)	Para controlar o preço de habitação entre zonas / cidades, são utilizadas 17 variáveis <i>dummy</i> <sup>d</sup>	Positivo (+) Negativo (-)
<b>Caraterísticas de vizinhança</b>					
<b>PIP (Finanças, E_S, C_M, HOSP)</b>	Perto instalação pública	Indica-nos se a habitação se encontra perto de escola, câmara municipal, hospital...	Limsombunch ai <i>et al.</i> (2004)	<i>Dummy</i> = 1 se a habitação se encontra perto de instalações públicas; 0 caso contrario	Positivo (+)

**Nota.** <sup>a</sup> Para a variável número de quartos (N\_Q), usamos 3 variáveis *dummy*, N\_Q1 (um quarto); N\_Q2 (dois quartos); N\_Q3 (três quartos). <sup>b</sup> Representa o número de casas de banho (NCB), usamos 2 variáveis *dummy*, NCB\_1 (uma casa de banho); NCB\_2 (doas casas de banho). <sup>c</sup> Representa o número de andares (ANDAR) das habitações onde usamos 3 variáveis *dummy*, ANDAR\_0 (rés do chão); ANDAR\_1 (primeiro andar); ANDAR\_2 (segundo andar). <sup>d</sup> Neste caso, usamos 17 variavas *dummy*, na determinação da localização de habitação: S\_M representa Santa Maria; AG\_SM representa Algodoeiro- Santa Maria; BN I\_ES representa Bairro Novo I- Espargos; BN II\_ES representa Bairro Novo II- Espargos; BN\_ES representa Bairro novo-Espargos; CM\_ES representa Chã de Matias- Espargos; ESM representa Entrada de Santa Maria; H\_ES representa Hortelão- Espargos; LSM representa Leste de Santa Maria; MC\_ES representa Morro curral- espargos; NSM representa Norte de Santa Maria; PL representa Palmeira; PP\_SM representa Ponta Preta- Santa Maria; PAS\_SM representa Praia Antónia Sousa- Santa Maria; PG\_ES representa Preguiça- Espargos; RF\_ES representa Ribeira Funda- Espargos; MD representa Murdeira.

### 3.3. Forma funcional utilizado no MPH

Não há nenhuma base teórica para a escolha de uma forma funcional específica para uma regressão hedônica. Vários autores testaram diferentes formas funcionais, tais como linear, semi-logarítmica, dupla logarítmica.

A pesquisa mostra, no entanto, que a escolha da forma funcional envolve um “*trade-off*” entre minimizar o erro de previsão e o problema de multicolinearidade (Weirick & Ingram, 1990).

Weirick e Ingram (1990) testaram a forma linear, semi-logarítmica e dupla logarítmica, num conjunto de dados compostos por 7.111 observações, relativamente a todas as vendas de residências unifamiliares no período de 1979 a 1987, na região centro-sul dos Estados Unidos precisamente em Golfo. Os resultados mostram a superioridade da forma semi-logarítmica sobre as outras formas funcionais. Para os autores, quando se trata da forma linear tanto as variáveis dependentes como as independentes possuem a especificação linear; na forma Semi-Logarítmica ou Semi-Log a variável dependente é transformada logaritmicamente, mantendo a linearidade das variáveis independentes; na forma dupla Logarítmica ou Log-Log tanto as variáveis dependentes como as independentes são transformações logarítmicas.

Bello e Moruf (2010) utilizaram o modelo de preço hedônico para estudar os preços das habitações em Lagos, na Nigéria. Também testaram três formas funcionais: forma linear, forma semi-logarítmica e forma dupla logarítmica. Das três formas testadas a forma semi-logarítmica é o que apresenta melhor ajuste, especialmente no que diz respeito ao coeficiente de determinação, ou seja, os resultados também mostraram a superioridade da forma semi-logarítmica sobre outras formas funcionais.

Thayer, Albers e Rahmatian (1992) optaram por usar duas formas funcionais: linear e semi-logarítmica. A principal razão desta escolha é o fato da grande parte da análise realizada requerer que a variável distância seja convertido em medidas de zero/um, onde dupla logarítmica e outras formas funcionais não podem ser usadas.

Em termos práticos, é sempre necessário experimentar várias formas funcionais como a forma cubica, quadrática, exponencial, logística entre outros, com o objetivo de verificar aquela que melhor consegue passar aos testes da regressão múltipla, apresentando melhor poder explicativo da variável independente.

No nosso estudo as formas lineares: semi-logarítmica, dupla logarítmica, quadrática e cubica foram testadas na tentativa de se obter a melhor forma funcional.

Equação do modelo de preços hedónicos:

$$\begin{aligned}
 Y_i = & \beta_0 + \beta_1 \cdot \text{Área}_i + \sum_{j=1}^2 \beta_{1+j} \cdot \text{NCB}_{j,i} + \sum_{j=1}^3 \beta_{3+j} \cdot \text{NQ}_{j,i} & (5) \\
 & + \sum_{j=1}^3 \beta_{6+j} \cdot \text{ANDAR} + \beta_{10} \cdot \text{VRD}_i + \beta_{11} \cdot \text{TRR}_i + \sum_{j=1}^{17} \beta_{11+j} \cdot \text{Loc}_{j,i} \\
 & + \beta_{29} \cdot \text{CM\_FN}_i + \beta_{30} \cdot \text{E\_S}_i + \beta_{31} \cdot \text{Hosp}_i + \varepsilon_i
 \end{aligned}$$

Onde:

$y_i$ - é o preço;

NCB - representa o número de casas de banho;

N\_Q - representa o número de quarto;

ANDAR- representa o número de andar;

VRD- representa a existência de varanda;

TRR- representa a existência de terraço;

Loc- representa a localização das habitações;

CM\_FN- habitações perto de camara municipal e fianças;

E\_S- habitações perto de escola secundária;

Hosp- habitações perto de hospital.

### **3.4. Justificação de utilização do modelo de RNA**

A RNA é uma das técnicas utilizada em situações de previsão e estimação (Berry & Linoff, 2004). Neste estudo, serão utilizadas as RNA para estimar o preço das habitações da Ilha do Sal, em Cabo Verde.

Devido à evolução, às características do mercado imobiliário, assim como à importância de avaliação do imóvel, existe a necessidade de implementar novas metodologias de avaliação, que atendam às especificidades deste tipo de avaliação. Na utilização de outros modelos de formação do preço do imóvel, podemos deparar-nos com problemas que podem reduzir a precisão das estimativas de valores, principalmente no que diz respeito ao conhecimento da forma funcional existente entre as variáveis e às dificuldades de estimação dos parâmetros referentes à distribuição espacial dos imóveis (Neto, 2006).

A escolha desta técnica como técnica de modelagem também foi sustentada em algumas vantagens elencadas na literatura, tais como:

- A capacidade e o poder de lidar com estrutura não lineares (Berry & Linoff, 2004);
- O uso de redes neuronais não requer um conhecimento a priori da relação entre as variáveis, já que elas são auto adaptáveis, ou seja são capazes de gerar modelos não lineares sem qualquer conhecimento prévio sobre a relação entre as variáveis de entrada e saída, podendo assim ser mais genéricas e flexíveis do que as metodologias tradicionais (Passari, 2003).
- A relação de dependência entre variáveis é detetada automaticamente pela rede, logo podem ser ignoradas as que não contribuem para a previsão (Passari, 2003).

### **3.5. Processo de estimação do preço das habitações pela rede**

Para estimar o preço das habitações através da rede, primeiramente é necessário a escolha dos sinais de entrada, ou seja, as informações de um determinado problema, estes sinais têm associado a eles um peso, denominada de peso sináptico (são números, positivos ou negativos, dando a força, ou eficácia, da conexões existente entre neurónios). Estes sinais representam a entrada na rede, onde são multiplicados pelos respetivos pesos, posteriormente é realizada o somatório da multiplicação ente as entradas e os pesos, assim existe efetivamente um neurónio. No valor dado como entrada, o neurónio poderá deixar passar ou inibir, em alguns dos casos a saída é alterada de acordo com o valor da entrada. Quando o sinal passa, logo aplica-se a função, que produz a transformação das entradas de cada neurónio (Tabela 2). A função ativação é aplicada consoante determinado problema, onde o resultado é a saída desejada.

### **3.6. Software utilizado e medidas de desempenho da rede**

Neste estudo irá ser utilizado o *software* de análise estatística SPSS.20. O uso do *software* SPSS facilita a estimação de rede neuronal, uma vez que o processo de estimação é realizado em varias etapas, e é preciso uma avaliação constante das soluções do modelo a fim de melhorar o grau de ajustamento.

O módulo de redes neuronais do SPSS dá-nos uma opção que permite selecionar automaticamente a melhor arquitetura da rede. O SPSS apresenta duas aplicações de redes neurais, o *perceptron* de múltiplas camadas (MLP) e a função de base radial (RBF), onde as redes são supervisionadas para que haja uma comparação dos resultados previstos pelo modelo com valores conhecidos das variáveis “alvo”.

A fim de avaliar o desempenho absoluto e relativo dos diversos modelos de previsão, foram desenvolvidas diversas medidas de desempenho. A aplicação das medidas de erro representa uma tentativa abrangente para avaliar e comparar o desempenho dos diferentes métodos aplicáveis para a avaliação. A escolha das medidas de desempenho não é arbitrária, ela deve ser feita, com base nas características de um problema (Passari, 2003).

Para avaliar o grau de ajustamento (desempenho) da rede, o SPSS calcula duas medidas: a soma dos quadrados de erro (SSE) e o erro relativo, em que a melhor rede, ou seja, a

mais eficaz, é a que apresenta menor erro relativo, num conjunto de teste (Teixeira, 2011). Na Equação 6 está descrita a fórmula do cálculo do erro relativo.

$$ER = \frac{\sum_{m=1}^m (y^m - \hat{y}^m)^2}{\sum_{m=1}^m (y^m - \bar{y}^m)^2} \quad (6)$$

Onde:

$y^m$  é o preço observado para caso m

$\hat{y}^m$  é a estimativa de preço para o caso m

$\bar{y}^m$  é media dos valores observados

m é o conjunto do caso de teste

Inmaculada Pulido-Calvoa, Roldán e Ruiz-Navarro (2007), basearam-se na seguinte fórmula do erro relativo para calcular a eficiência da rede.

$$ER = 1 - \frac{\sum_{m=1}^m (y^m - \hat{y}^m)^2}{\sum_{m=1}^m (y^m - \bar{y}^m)^2} \quad (7)$$

A soma dos quadrados dos erros (SSE) é o total dos erros ao quadrado verificado.

Equação da soma dos quadrados dos erros (SSE):

$$SSE = \sum_{m=1}^m (y^m - \hat{y}^m)^2 \quad (8)$$

Para avaliar o grau de ajustamento (desempenho) do modelo de preço hedónico, o SPSS calcula o  $R^2$ .

$$R^2 = \frac{\left( M \sum_{m=1}^m y^m \hat{y}^m - \sum_{m=1}^m y^m \sum_{m=1}^m \hat{y}^m \right)^2}{\left( M \sum_{m=1}^m y^{(m)^2} - \left( \sum_{m=1}^m y^{(m)} \right)^2 \right) \left( M \sum_{m=1}^m \hat{y}^{(m)^2} - \left( \sum_{m=1}^m \hat{y}^{(m)} \right)^2 \right)} \quad (9)$$

Estas estatísticas seriam suficientes para comparar as redes entre si, mas como o nosso objetivo é comparar o modelo de redes neuronais e o modelo do preço hedónico, logo será necessário calcular outras medidas.

Os autores Mccluskey *et al.* (2013), Peddy (2011) e Selim (2009) utilizaram várias medidas, entre elas o erro médio percentual absoluto (MAPE), o erro quadrado médio (MSE), o erro absoluto médio (MAE) e a raiz quadrada do erro quadrado médio (RMSE), para avaliar a capacidade de generalização da rede, medida pelo ajustamento de um conjunto de teste.

O erro médio percentual absoluto (MAPE) é uma medida em termos percentuais, para que se tenha uma visão do erro comparado com o valor previsto.

$$MAPE = \frac{1}{M} \sum_{m=1}^m \left| \frac{y^m - \hat{y}^m}{y^m} \right| \quad (10)$$

O erro quadrado médio (MSE) é uma medida que calcula a média dos erros observados entre o valor gerado pela rede neuronal, isto é, a previsão da rede e o valor desejado (o valor observado na prática).

$$MSE = \frac{1}{M} \sum_{m=1}^m (y^m - \hat{y}^m)^2 \quad (11)$$

O erro absoluto médio (MAE) é o erro médio tomado em termos absolutos, para que um erro positivo não seja anulado por outro negativo.

$$MAE = \frac{1}{M} \sum_{m=1}^m |y^m - \hat{y}^m| \quad (12)$$

A raiz quadrada do erro quadrado médio (RMSE) é raiz quadrada da média dos valores ao quadrado da previsão de erros da rede.

$$RMSE = \sqrt{\frac{1}{M} \sum_{m=1}^m (y^m - \hat{y}^m)^2} \quad (13)$$

As metodologias das especificações do modelo e arquitetura são difíceis de comparar diretamente, as medidas selecionadas são para testar como diferentes especificações e arquiteturas de modelo podem ser comparados em termos de precisão e previsão. O modelo com menor número de erro é mais eficiente.

## 4. APRESENTAÇÃO E DISCUSSÃO DE RESULTADOS

Esta secção descreve e analisa os resultados, começando pela apresentação das estatísticas descritivas. Em seguida apresentam-se os resultados da rede neuronal artificial; seguindo-se a apresentação do estudo através do modelo de preço hedónico, finalmente, efetuamos uma comparação entre os dois modelos.

### 4.1. Estatística descritiva

A amostra é representada por 1092 observações. As Tabelas 7 e 8 apresentam a análise estatística da variável dependente e da variável independente quantitativa. Na amostra o valor médio da variável “preço de venda” foi de 6.509.598,26 escudos cabo-verdianos (CVE) equivalente a 59.178,166 euros, com um desvio padrão de 2.783.962,269 CVE. O valor mínimo e o valor máximo observados são representados por 25.000,00 CVE e 13.732.595,00 CVE, respetivamente a dimensão (Área/ m<sup>2</sup>) das habitações tem um valor médio de 97,44 m<sup>2</sup> e um desvio padrão de 39,747 m<sup>2</sup>, o mínimo e o máximo valor observados são 20,00 m<sup>2</sup> e 256,00 m<sup>2</sup>.

Tabela 7: Resumo estatístico da variável dependente e variável independente continua

	N	Média	Desvio padrão	Mínimo	Máximo
<b>Variável Dependente</b>					
<b>Preço de venda</b>	1092	6.509.598,26	2.783.962,269	25.000,00	13.732.595,00
<b>Variável Independente</b>					
<b>Área / (m<sup>2</sup>)</b>	1092	97,44	39,747	20,00	256,00

Na Tabela 8 está exposta a frequência e o valor percentual de cada variável independente qualitativa da amostra. Podemos observar que 51,6% das habitações da Ilha do Sal possuem varanda, e 34,7% têm terraço. Constatamos que cerca de 49,3% das habitações são constituídas por dois quartos e 66,8% têm uma casa de banho, e 31,7% encontram-se situadas no primeiro andar. A maioria das habitações da Ilha do Sal (83%) encontram-se localizadas na cidade de Santa Maria, e cerca de 84,3% localizam-se perto da câmara

municipal e das finanças. Os restantes 1,5% localizam-se perto da escola secundária, e 4% perto do hospital.

Tabela 8: Resumo estatístico das variáveis independentes (dummy)

<b>Caraterísticas Estruturais</b>	<b>N</b>	<b>%</b>
Número de varandas	564	51,6%
Número de terraços	379	34,7%
<b>Número de quartos</b>		
Quarto 1	424	38,8%
Quarto 2	538	49,3%
Quarto 3	117	10,7%
<b>Número de casas de banho</b>		
Casa de banho 1	730	66,8%
Casa de banho 2	330	30,2%
<b>Andar</b>		
Rés-do-chão	229	21,0%
Primeiro andar	346	31,7%
Segundo andar	334	30,6%
<b>Caraterísticas de localização</b>	<b>N</b>	<b>%</b>
<b>Cidades/zonas</b>		
Santa Maria	906	83,0%
Algodoeiro – Santa Maria	4	0,4%
Bairro novo – espargos	10	0,9%
Bairro novo I - Espargos	2	0,2%
Bairro novo II - Espargos	9	0,8%
Chã de Matias - espargos	4	0,4%
Entrada de Santa Maria	24	2,2%
Hortelão - Espargos	8	0,7%
Leste de Santa Maria	4	0,4%
Morro Curral – Espargos	2	0,2%
Norte de Santa Maria	16	1,5%
Palmeira	8	0,7%
Ponta Preta – Santa Maria	2	0,2%
Praia António Sousa – Santa Maria	45	4,1%
Preguiça- Espargos	4	0,4%
Ribeira Funda	5	0,5%
Murdeira	28	2,6%

Tabela 8: **Resumo estatístico das variáveis independentes (dummy) (cont.)**

<b>Caraterísticas de vizinhança</b>	<b>N</b>	<b>%</b>
<b>PIP</b>		
Câmara Municipal/ Finanças	921	84,3%
Hospital	4	0,4%
Escola Secundaria	16	1.5%

## **4.2. Estimação do preço das habitações através de RNA**

Para a realização deste estudo foi utilizado o SPSS, versão 20, este *software* simula uma rede de múltiplas camadas do tipo *feedforward*, com três tipos de função ativação, a tangente hiperbólica, sigmoide e identidade. Não foram efetuadas nenhuma escolhas referente a arquitetura da rede, uma vez que o *software* seleciona automaticamente o desenho e as principais características da rede. Foram usadas 30% da amostra para o teste e 70% para o treinamento, tal como Amaro (2012). Estas observações foram escolhidas de forma aleatória.

O nosso processo iniciará com a seleção das variáveis independentes da camada de entrada. É habitual existir um conjunto de variáveis explicativas, mas não é conveniente usá-las todas, porque algumas têm pouco poder explicativo (Teixeira, 2011).

Assim, na tentativa de obter o melhor desempenho da rede, contruímos 6 cenários de teste, para verificar o comportamento da rede com a introdução de novas variáveis explicativas.

Os nossos cenários são constituídos por:

**Cenário 1** - Área (variável quantitativa) + localização (variável qualitativa);

**Cenário 2** - Área + localização + número de quartos + número de casas de banho (variáveis qualitativas);

**Cenário 3** - Área + localização + número de quartos + número de casas de banho + existência de varanda + existência de terraço (variáveis qualitativas);

**Cenário 4** - Área + localização + número de quartos + números de casas de banho + existência de varanda + existência de terraço + andar (variável qualitativa);

**Cenário 5** - Área + localização + número de quarto + número de casas de banho + existência de varanda + existência de terraço + andar + perto instituições públicas (camara municipal e fianças, hospital, escola secundaria) (variáveis qualitativas);

**Cenário 6** - Área + localização + número de quarto + número de casas de banho + existência de varanda + existência de terraço + andar + perto instituições públicas (camara municipal e fianças, hospital, escola secundária) + zona (variável qualitativa);

Este último cenário foi construído na tentativa de melhorar ainda mais a nossa rede. De maneira a minimizar o valor do erro relativo, então criamos a variável zona, onde agrupamos todas as zonas localizada em Santa Maria em que denominamos de zona 1 e as restantes zonas foram denominadas de zona 2.

Tabela 9: **Resumo dos resultados da experiencia por cenário**

Cenários	Valor do erro relativo
Cenário 1	0.587
Cenário 2	0.477
Cenário 3	0.512
Cenário 4	0.519
Cenário 5	0.485
Cenário 6	<b>0.476</b>

**Nota.** O erro relativo foi calculado baseando-se na seguinte fórmula:

$$ER = \frac{\sum_{m=1}^m (y^m - \hat{y}^m)^2}{\sum_{m=1}^m (y^m - \bar{y}^m)^2}$$

Através da Tabela 9 verifica-se que a melhor rede elaborada pelo *software*, ou seja, a que levou ao menor erro relativo, foi a que estava suportada no cenário 6.

Na Tabela 10 encontram-se os resultados do erro relativo do treinamento e do teste do melhor cenário (cenário 6), onde constatamos que o erro relativo do treinamento (0,373) é inferior ao erro relativo dos testes (0.476), ou seja, 0,103.

Tabela 10: **Resumo do modelo, em termo de medida de desempenho**

<b>Treinamento</b>	Soma dos erros quadráticos	141,251
	Erro relativo	0,373
	Regra de parada usada	1 etapa (s) consecutiva (s) sem nenhuma diminuição do erro
	Tempo de treinamento	0:00:01.48
<b>Teste</b>	Soma dos erros quadráticos	76,926
	Erro relativo	0.476

Fonte: Elaborado pelo SPSS.

**Nota.** O erro relativo foi calculado baseando-se na seguinte fórmula:

$$ER = \frac{\sum_{m=1}^m (y^m - \hat{y}^m)^2}{\sum_{m=1}^m (y^m - \bar{y}^m)^2}$$

A Tabela 11 ilustra as informações principais da melhor rede neuronal (cenário 6), assim como as variáveis utilizadas, função de ativação e número de unidades de camada oculta.

Tabela 11: **informações da rede neuronal**

<b>Camada de entrada</b>	<b>Fatores</b>	1	Loc
		2	N_Q
		3	N_WC
		4	VRD
		5	TRR
		6	ANDAR
		7	CM_FN
		8	E_S
		9	Hosp
		10	ZONA
	<b>covariáveis</b>	1	Área
<b>Camada ocultas</b>	Número de unidades		42
	Método de reescalonamento para covariáveis		Padronizado
	Número de camadas ocultas		1
<b>Camada de saída</b>	Número de unidades na camada oculta		9
	Função ativação		Tangente hiperbólica
	Variáveis dependentes 1		Preço_ venda

Tabela 11: **informações da rede neuronal** (cont.)

<b>Número de unidades</b>	1
Método de reescalonamento para dependentes de escala	Padronizado
Função de ativação	Identidade
Função de erro	Soma dos quadrados

Fonte: Elaborado pelo SPSS.

**Nota.** A Loc representa a localização, N\_Q representa o número de quarto, NCB representa o número de casas de banho, VRD representa a existência de varanda, TRR representa a existência de terraço, ANDAR representa o andar, CM\_FN, E\_S, Hosp representa perto instituições de camara municipal e fianças, hospital, escola secundária.

No Anexo 1, podemos ver a estimativa de rede neuronal, assim como os pesos sinápticos e os seus respetivos sinais.

A Tabela 12 indica-nos a importância relativa e normalizada, ou seja, a capacidade explicativa de cada variável independente na estimação do preço das habitações na Ilha do Sal. Podemos verificar que a variável com maior poder explicativo no preço é a área, com uma importância relativa de 38,7%, em seguida é a localização, com uma importância relativa de 24,4%. As variáveis que apresentam menor poder explicativo no preço do imóvel é a variável perto de câmara municipal e das finanças com uma importância relativa de 1,6%, seguida da variável existência de varanda com uma importância relativa de 1,8%. Estas variáveis com menos importância, se forem excluídas do modelo, tornarão os resultados finais menos realistas (Amaro, 2012).

Tabela 12: **Importância das variáveis independentes**

	<b>Importância</b>	<b>Importância normalizada</b>
Loc	0,244	63,1%
N_Q	0,069	17,9%
NCB	0,030	7,8%
VRD	0,018	4,7%
TRR	0,036	9,3%
ANDAR	0,049	12,7%
CM_FN	0,016	4,2%
E_S	0,028	7,2%
Hosp	0,046	11,9%
ZONA	0,075	19,4%
Área	0,387	100%

Tabela 12: **Importância das variáveis independentes** (cont.)

Fonte: Elaborado pelo SPSS

**Nota.** A Loc representa a localização, N\_Q representa o número de quarto, NCB representa o número de casas de banho, VRD representa a existência de varanda, TRR representa a existência de terraço, ANDAR representa o andar, CM\_FN, E\_S, Hosp representa perto instituições de camara municipal e fianças, hospital, escola secundária.

A Figura 2, ilustra a capacidade explicativa das variáveis da melhor rede, sustentada no cenário 6.

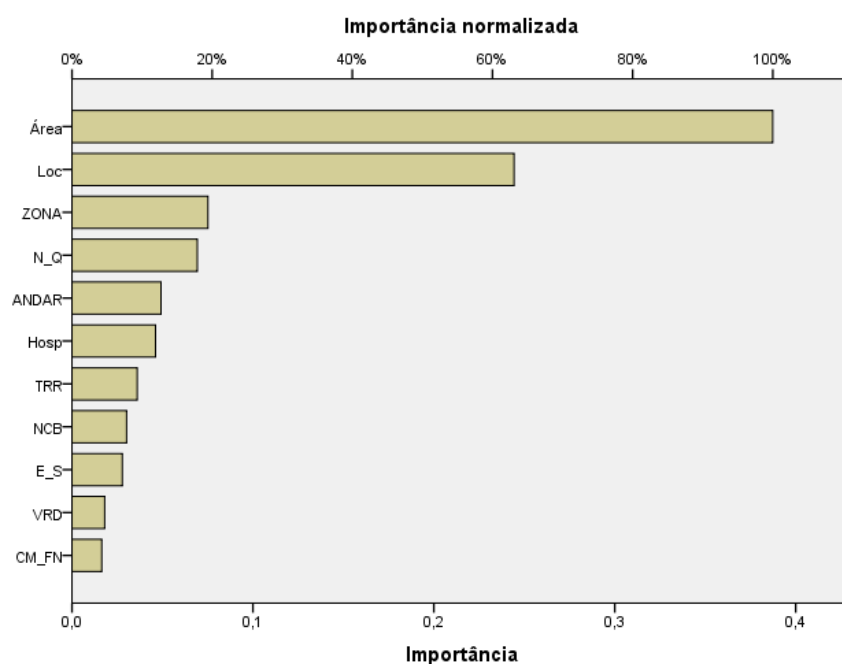


Figura 2: **capacidades explicativas das variáveis independentes**

Fonte: Elaborado pelo SPSS

**Nota.** A Loc representa a localização, N\_Q representa o número de quarto, NCB representa o número de casas de banho, VRD representa a existência de varanda, TRR representa a existência de terraço, ANDAR representa o andar, CM\_FN, E\_S, Hosp representa perto instituições de camara municipal e fianças, hospital, escola secundária.

A Figura 3, apresenta o ajuste do preço das habitações, em que verificamos que a maioria das observações está concentrada no valor de 7.500.000,00 (valor aproximado) de escudo cabo-verdiano (CVE), equivalente a 68.181,81 €.

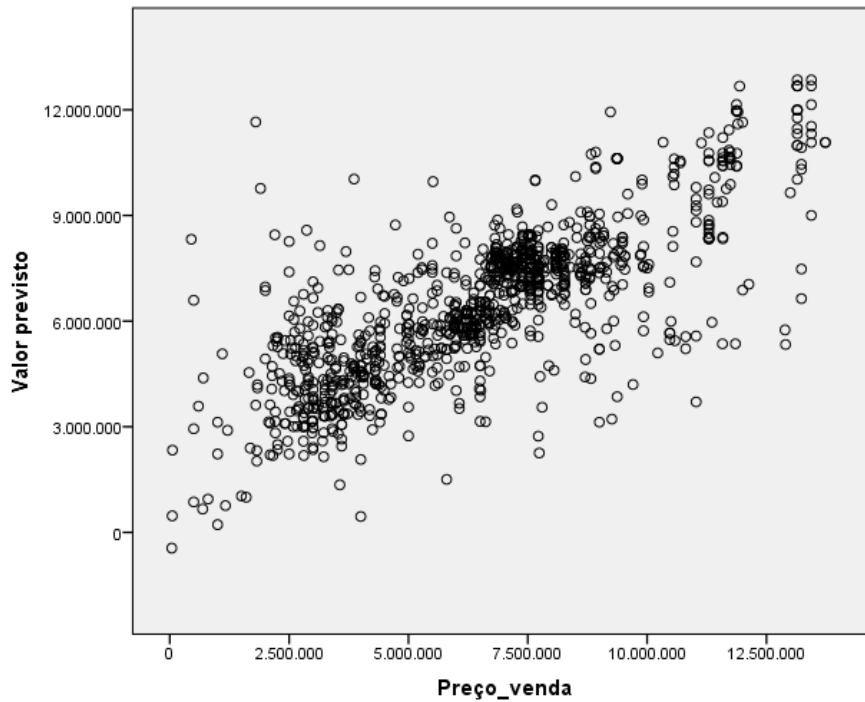


Figura 3. Preço observado Vs. Preço estimado pela rede neuronal

Fonte: Elaborado pelo SPSS

Agora vamos observar o funcionamento da nossa rede, como já foi referido anteriormente, ela é constituída pela camada de entrada, de saída e de uma camada oculta. Os valores da entrada da variável independente qualitativa (área) foram padronizadas, e as variáveis independentes qualitativas (localização, número de quartos, número de casas de banho e andar) não sofreram alterações. As variáveis como existência de varanda, existência de terraço, perto de câmara municipal e das finanças, escola secundária e hospital foram introduzidos na rede através de variáveis *dummy*.

A padronização das variáveis é efetuada de acordo com a Equação 14:

$$\frac{E - \bar{E}}{S} \quad (14)$$

Onde:

$E$ - representa o valor da observação;

$\bar{E}$ - é a media do valore observado;

S- é o desvio padrão.

Estes valores entram na rede através da camada de entradas, são multiplicadas pelo peso da conexão entre os neurónios da camada oculta, e por fim efetua-se a soma dos produtos como mostra a Equação 15, o valor de propensão (*bias*) também é adicionado.

$$u_1 = w_{01} * E_0 + w_{11} * E_1 + w_{21} * E_2 \dots w_i * E_i + bias \quad (15)$$

Onde:

$E_i$  - são as variáveis da entrada;

$w_i$  - são as pesos de conexões.

Os pesos das conexões  $w_i$  são calculados pelo *software* (Anexo 2).

O resultado da Equação 15 é aplicado na função ativação (tangente hiperbólica), que resulta posteriormente em valores do neurónio da camada oculta. O processo para obter o valor da saída do neurónio é o mesmo aplicado para obter o valor da camada oculta a partir da camada de entrada, a diferença está no uso da função de ativação, que neste caso é a identidade, e no peso de conexão existente entre camada oculta e camada de saída (Anexo 2). Se no caso o resultado do valor da saída for padronizado, logo teremos que converte-lo para se obter o valor desejado, esta conversão é efetuada com base na equação de padronização dos valores, como mostra a Equação 16.

$$S_i * S + \bar{E} \quad (16)$$

Onde:

$S_i$  - Valor da saída;

$S$  - Desvio padrão;

$\bar{E}$  - Média do valor observado.

Com o conhecimento dos pesos de conexões entre as camadas e as funções de ativação, pode-se calcular cada entrada na rede, que possibilitará a avaliação da sua contribuição na formação do preço das habitações.

Vamos agora dar um exemplo concreto, a semelhança do estudo de Teixeira (2011), iremos usar o caso que possui menor erro observado. Trata-se de uma habitação constituída por três quartos, duas casas de banho, com 163 m<sup>2</sup> de superfície, localizada em Santa Maria, situado no segundo andar, possui uma varanda e um terraço e fica perto da câmara municipal e das finanças, estas são as características que constituem a entrada na rede. O valor da entrada é multiplicado pelo respetivo peso, o mesmo é produzido pela própria rede, posteriormente é realizado o somatório da multiplicação e aplicado a função ativação (tangente hiperbólica). O processo para obter o valor da saída, é o mesmo, a diferença está no uso da função de ativação (identidade), e no peso de conexão existente entre a camada oculta e a camada de saída (Anexo 2). O valor do preço observado é de 9.893.761,00 escudos, o valor estimado pela rede é de 9.893.034,00 escudos. Verificamos que o erro entre o valor do preço real da venda e o preço estimado pela rede é apenas de 727,4 escudos, equivalente a 6,612 euros.

### **4.3. Estimação do preço das habitações através do MPH**

Para este estudo também foi utilizado o *software* SPSS 20, onde usamos o modelo da regressão linear múltipla na estimação do preço de habitação na Ilha do Sal em Cabo Verde. A estimação do modelo do preço hedónico baseia-se nos valores reais das vendas de 1092 observações.

Para a realização do estudo é necessário a escolha de uma forma funcional. A literatura não determina uma forma funcional específica, portanto os modelos diferem de acordo com os dados e conjuntos de realidade que será estudado. Algumas das formas como: forma linear, quadrática, cubica, dupla logarítmica e semi-logarítmica foram testadas, mas a que apresentou melhor ajuste foi semi-logarítmica, com um coeficiente de determinação de 53,6%, as restantes formas apresentaram um coeficiente de determinação menor.

Tal como Bello e Moruf (2010) e Weirick e Ingram (1990), vamos utilizar a especificável funcional semi-logarítmica. Esta especificável transforma logaritmicamente a variável

dependente (preço), mantendo a linearidade das variáveis independentes. A nossa equação hedónica é definida de acordo com a equação 5.

Nesta forma funcional o coeficiente  $\beta$  representa a variação relativa, ou seja, uma variação percentual no preço da habitação em consequência da variação em unidade das variáveis explicativas.

Uma vez que utilizamos o modelo da regressão linear, para a realização da presente análise, é necessário verificar a existência de multicolinearidade entre as variáveis independentes. Se no caso existir multicolinearidade, a estimação dos parâmetros do modelo poderá ser afetado, o que impossibilitará a utilização de todas as variáveis.

Uma forma de analisar estes problemas é através da análise do *Variance Inflation Factor* (VIF), Se o valor do VIF for superior a 10 significa que existem problemas de multicolinearidade. Uma outra forma de analisar a multicolinearidade é através da análise do coeficiente de correlação, particularmente pela análise da correlação de Pearson. Esta análise calcula o grau de correlação entre duas variáveis, numa escala que varia entre -1 e 1. Para evitar o problema de multicolinearidade, recomenda-se que o seu valor em módulo não exceda a 0,75.

Nesta análise, consideram-se a correlação entre variáveis através do método *Variance Inflation Factor* (VIF).

Tabela 13: Correlação entre as variáveis

Variáveis	Tolerância	VIF
Área	0,767	1,304
S_M	0,021	47,199
AG_SM	0,731	1,368
BN_ES	0,501	1,997
BNI_ES	0,832	1,203
BNII_ES	0,300	3,332
CM_ES	0,147	6,817
ESM	0,310	3,222
LSM	0,717	1,394
H_ES	0,562	1,780
MC_ES	0,839	1,191
NSM	0,410	2,437
PL	0,115	8,686
PP_SM	0,841	1,189
PAS_SM	0,200	4,992
RF_ES	0,122	8,226
MD	0,280	3,568
NQ_1	0,100	9,675
NQ_2	0,097	9,685
NQ_3	0,103	9,665
NCB_1	0,120	8,331
NCB_2	0,121	8,283
ANDAR_0	0,508	1,968
ANDAR_1	0,491	2,035
ANDAR_2	0,494	2,023
VRD	0,654	1,530
TRR	0,597	1,675
CM_FN	0,032	30,908
Hosp	0,726	1,378
E_S	0,047	21,459

**Nota.** A Loc representa a localização; N\_Q representa o número de quarto; NCB representa o número de casas de banho; VRD representa a existência de varanda; TRR representa a existência de terraço; ANDAR representa o andar; CM\_FN, E\_S, Hosp representa perto instituições de camara municipal e fianças, hospital, escola secundária; S\_M representa Santa Maria; AG\_SM representa Algodoeiro- Santa Maria; BN I\_ES representa Bairro Novo I- Espargos; BNII\_ES representa Bairro Novo II- Espargos; BN\_ES representa Bairro novo-Espargos; CM\_ES representa Chã de

Tabela 13: **correlação entre as variáveis** (cont.)

---

Matias- Espargos; ESM representa Entrada de Santa Maria; H\_ES representa Hortelão- Espargos; LSM representa Leste de Santa Maria; MC\_ES representa Morro curral- espargos; NSM representa Norte de Santa Maria; PL representa Palmeira; PP\_SM representa Ponta Preta- Santa Maria; PAS\_SM representa Praia Antónia Sousa- Santa Maria; PG\_ES representa Preguiça- Espargos; RF\_ES representa Ribeira Funda- Espargos; MD representa Murdeira.

A maioria das variáveis apresenta o VIF inferiores ao valor de referência: 10, então conclui-se que não há nenhuma multicolinearidade para essas variáveis. No entanto, os valores VIF são mais de 10 para as variáveis, perto de câmara municipal e das finanças, perto de escola secundária e das habitações localizadas em Santa Maria, estas são responsáveis pela existência de multicolinearidade. Esta situação pode estar relacionada com o fato dessas variáveis (perto da camara municipal e das finanças e perto da escola secundaria) encontrarem-se localizadas em Santa Maria.

Agora vamos dar inicio a construção do modelo da regressão. O nosso modelo inclui todas as variáveis independentes, exceto a variável Preguiça- Espargos (PG\_ES), provavelmente esta exclusão deu-se na tentativa de melhorar a linearidade dos dados e consequentemente melhorar o nosso coeficiente de determinação ( $R^2$ ).

O grau de ajuste do modelo é de 53,6% e o seu valor ajustado é de 52,2% uma diferença de 1.4%. Isto significa que aproximadamente 53,6% da variação que ocorre no preço é explicada pelas variáveis independentes incluídas no modelo. Comparando com outros estudos, como por exemplo de Teixeira (2011) e Carreira (2011), o nosso coeficiente de determinação é bastante baixo.

Na observação do gráfico de dispersão dos pontos do preço observado e do preço estimado, verificamos que o ajuste do modelo é razoável.

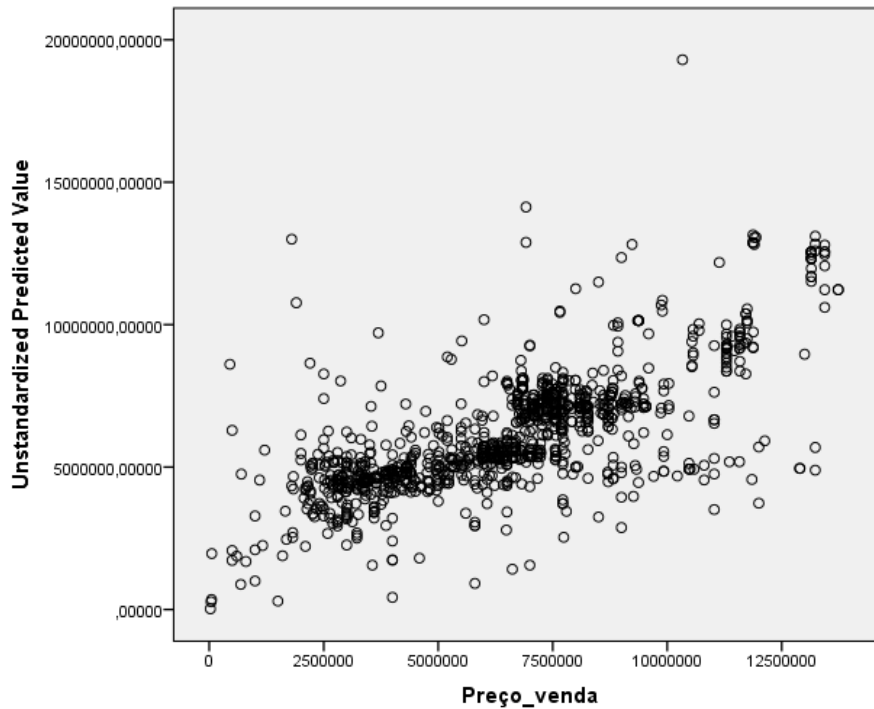


Figura 4. Modelo do preço hedónico ajustado

O teste F (validade dos ajustamentos) Este teste de hipóteses permite avaliar, a partir de estimativas amostrais se, de facto, algumas das variáveis independentes podem ou não influenciar a variável dependente, isto é, se o modelo ajustado é ou não significativo.

Hipóteses:

$$H_0: \beta_1 = \beta_2 \dots = \beta_k = 0$$

$$H_1: \text{Existe pelo menos uma } \beta_i \neq 0$$

Como o nível de significância de 5% (0,05) é maior que o *p-value* (0,000), logo rejeita-se a hipótese  $H_0$ , assim, há evidências estatísticas suficientes para afirmar que existe pelo menos uma variável independente significativa para o modelo.

Tabela 14: coeficiente das variáveis e o nível de significância

Variáveis	Influência prevista	$\beta_i$	t
Constante		14,266	(20,086)
Área	+	0,07***	(18,703)
S_M	+/-	0,207	(0,822)
AG_SM	+/-	0,793***	(2,981)
BN_SM	+/-	-1,643***	(-8,058)
BN I_ES	+/-	-0,397	(-1,126)
BN II_ES	+/-	-0,333	(-1,199)
CM_ES	+/-	-2,368***	(-3,986)
ESM	+/-	0,453***	(2,693)
H_ES	+/-	-1,623***	(-7,545)
LSM	+/-	0,188	(0,700)
MC_ES	+/-	-0,391	(-1,114)
NSM	+/-	-0,431**	(-2,411)
PL	+/-	-1,776***	(-3,737)
PP_SM	+/-	0,249	(0,709)
PAS_SM	+/-	-0,208	(-1,343)
RF_ES	+/-	-1,638***	(-2,805)
MD	+/-	-0,227	(-1,384)
NQ_1	+	0,383***	(2,907)
NQ_2	+	0,387***	(2,927)
NQ_3	+	0,460***	(3,326)
NCB_1	+	0,321***	(3,812)
NCB_2	+	0,360***	(4,182)
ANDAR_0	+	0,087	(1,844)
ANDAR_1	+	0,014	(0,334)
ANDAR_2	+	0,024	(0,559)
VRD	+	0,025	(0,726)
TRR	+	0,032	(0,861)
CM_FN	+	-0,174	(-0,828)
Hosp	+	-2,868***	(-10,738)
E_S	+	0,957*	(1,805)
$R^2$	<b>0,536</b>		
$R^2$ Ajustada	0,522		
F teste	<b>35,769***</b>		
Observações	1092		

Tabela 14: **coeficiente das variáveis e o nível de significância** (cont.)

**Nota:**  $\beta_i$  é a estimativa de coeficiente para variável  $i$ . Os níveis de significância são indicados por \*, \*\* e \*\*\* que representam os níveis de 10%, 5%, 1%. A coluna de influência indica que a variável correspondente é previsto ter + (positivo), - (negativo) +/- (varia de caso para caso), sobre o preço de venda. A Loc representa a localização; N\_Q representa o número de quarto; NCB representa o número de casas de banho; VRD representa a existência de varanda; TRR representa a existência de terraço; ANDAR representa o andar; CM\_FN, E\_S, Hosp representa perto instituições de camara municipal e fianças, hospital, escola secundária; S\_M representa Santa Maria; AG\_SM representa Algodoeiro- Santa Maria; BN I\_ES representa Bairro Novo I- Espargos; BN II\_ES representa Bairro Novo II- Espargos; BN\_ES representa Bairro novo-Espargos; CM\_ES representa Chã de Matias- Espargos; ESM representa Entrada de Santa Maria; H\_ES representa Hortelão- Espargos; LSM representa Leste de Santa Maria; MC\_ES representa Morro curral- espargos; NSM representa Norte de Santa Maria; PL representa Palmeira; PP\_SM representa Ponta Preta- Santa Maria; PAS\_SM representa Praia Antónia Sousa- Santa Maria; PG\_ES representa Preguiça- Espargos; RF\_ES representa Ribeira Funda- Espargos; MD representa Murdeira.

As variáveis cujo coeficiente estimado tem valor positivo, significa que contribuem positivamente no aumento do preço das habitações, e as estimativas negativas produzem efeito contrário. A presença das características faz com que haja uma variação percentual no preço. Por exemplo uma habitação com dois quartos provoca uma variação percentual positivo no preço de 0,387%. Para a variável perto do hospital, provoca uma variação percentual negativa no preço de -2,868%.

Na Tabela 15, apresentamos o resumo do estudo empírico.

Tabela 15: **Resumo dos resultados empíricos**

Variáveis	Nome de variáveis	Hipótese a testar	Resultado empírico
<b>Características estruturais</b>			
<b>Área</b>	Tamanho da habitação	Positivo (+)	Sim
<b>N_Q</b>	Número de quartos	Positivo (+)	Sim
<b>NCB</b>	Número de casa de banho	Positivo (+)	Sim
<b>TRR</b>	Terraço	Positivo (+)	Não
<b>VRR</b>	Varanda	Positivo (+)	Não
<b>ANDAR</b>	Andar	Positivo (+)	Não

Tabela 15: **Resumo dos resultados empíricos** (cont.)

Variáveis	Nome de variáveis	Hipótese a testar	Resultado empírico
<b>Características de localização</b>			
<b>Loc</b>	Localização	Positivo (+) /negativo (-)	Sim: AG_SA, ESM, BN_ES, CM_ES, H_ES, PL, NSM, RF_ES  Não: S_M, B_Novo II, B_Novo I, MC_ES, LSM, PAS_SM, MD, PP_SM  Não integra no modelo: PG_ES
<b>Caraterísticas de vizinhanças</b>			
<b>PIP</b>	Perto instalação pública	Positivo (+)	Sim: E_S, Não: CM_FN, Hosp

**Nota.** Os resultados empíricos são identificados como “sim” indicam que as hipóteses foram corroboradas, e os resultados como “não” indicam que as hipóteses foram corroboradas. A Loc representa a localização; N\_Q representa o número de quarto; NCB representa o número de casas de banho; VRD representa a existência de varanda; TRR representa a existência de terraço; ANDAR representa o andar; CM\_FN, E\_S, Hosp representa perto instituições de camara municipal e fianças, hospital, escola secundária; S\_M representa Santa Maria; AG\_SM representa Algodoeiro-Santa Maria; BN I\_ES representa Bairro Novo I- Espargos; BN II\_ES representa Bairro Novo II- Espargos; BN\_ES representa Bairro novo-Espargos; CM\_ES representa Chã de Matias- Espargos; ESM representa Entrada de Santa Maria; H\_ES representa Hortelão- Espargos; LSM representa Leste de Santa Maria; MC\_ES representa Morro curral-Espargos; NSM representa Norte de Santa Maria; PL representa Palmeira; PP\_SM representa Ponta Preta- Santa Maria; PAS\_SM representa Praia Antónia Sousa- Santa Maria; PG\_ES representa Preguiça- Espargos; RF\_ES representa Ribeira Funda- Espargos; MD representa Murdeira.

Através da análise dos resultados obtidos na Tabela 15, pode constatar-se que as variáveis representativas na estimação do preço das habitações na Ilha do Sal, apresentam os níveis de significância de 10%, 5%, 1%. Assim, existe evidência empírica significativa para afirmar que o preço das habitações na Ilha do Sal é fortemente influenciado pela variável área, número de quartos, número de casas de banho, algumas variáveis de localização (Algodoeiro- Santa Maria, Bairro novo-Espargos, Chã de Matias- Espargos, Entrada de Santa Maria, Hortelão- Espargos, Norte de Santa Maria, Palmeira, Ribeira Funda-Espargos), e proximidade de escola secundária. Corrobora-se então essas hipóteses. Estes resultados estão de acordo com a evidência empírica apresentada por Carreira (2011), Limsombunchai *et al.* (2004) e Mccluskey *et al.*( 2013). Em suma, podemos dizer que encontramos resultados consistentes.

A equação do modelo do preço hedónico como mostra a equação 17, é obtida tendo por base os elementos da Tabela 13, permite estimar o preço das habitações na Ilha do Sal:

$$\begin{aligned}
\text{Preço} = & 14,266 + 0,07 \cdot \text{Área} + 0,321 + 0,360 \cdot \text{NCB} + 0,383 + 0,387 & (17) \\
& + 0,460 \cdot \text{NQ} + 0,087 + 0,014 + 0,024 \cdot \text{ANDAR} + 0,025 \cdot \text{VRD} \\
& + 0,032 \cdot \text{TRR} + 0,207 + 0,793 + (-1,643) + (-0,397) + (-0,333) \\
& + (-2,368) + 0,453 + (-1,623) + 0,188 + (-0,391) + (-0,431) \\
& + (-1,776) + 0,249 + (-0,208) + (-1,638) + (-0,227) \cdot \text{Loc} \\
& + (-0,174) \cdot \text{CM}_{FN} + 0,957 \cdot \text{E}_{S} + (-2,868) \cdot \text{Hosp}
\end{aligned}$$

Agora vamos seguir com um exemplo concreto, onde iremos usar o mesmo caso, utilizado no modelo de redes neuronais artificiais. Esta observação trata-se de uma habitação constituída por três quartos, duas casas de banho, com 163 m<sup>2</sup> de superfície, localizada em Santa Maria, situado no segundo andar, possui uma varanda e um terraço e fica perto da câmara municipal e das finanças. Onde o preço real da venda é de 9.893.761,00 escudos, e tem um preço estimado pelo MPH de 10852587,64 escudos. Verificamos que o erro entre o valor real da venda e o valor do preço estimado pelo MPH é de -958.826,6355 escudos, equivalente a -8716,60 euros. Este resultado mostra que neste caso o MPH tem a capacidade de estimar valores superiores em relação ao preço real da venda.

#### 4.4. Comparação de desempenho do modelo de RNA e MPH

A figura 5 mostra a comparação dos erros estimados entre os modelos. Podemos ver que, com exceção de alguns casos em que o erro é superior a 10.000.000,00 de escudos, equivalente a 90.909,09 euros, em termos absolutos, em todos os outros casos a magnitude do erro foi encontrada em torno de 5.000.000,00 escudos, equivalente a 45.454,54 euros, comparando com o estudo de Teixeira (2011) o resultado obtido é significativamente maior.

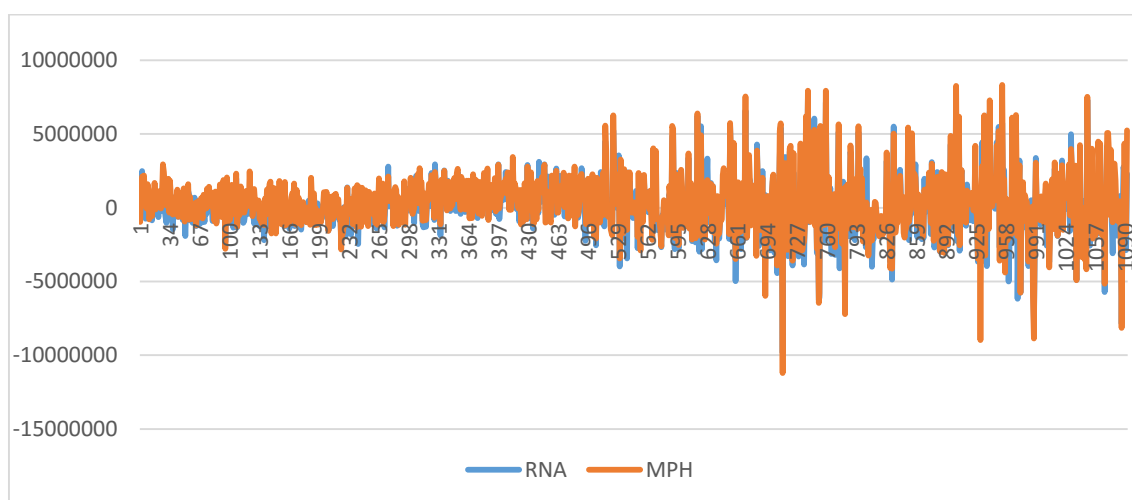


Figura 5. Erro do modelo estimado

Constatamos que no modelo de RNA, a percentagem do erro positivo é de 51,1% e do erro negativo é de 48,9%, uma diferença de 2,2%. O erro positivo médio é igual a 1.215.861,459 escudos, equivalente a 11.053,285,00 euros e o erro negativo médio é igual a -1.334.359,43 escudos, equivalente a -10.858,591 euros. Podemos ver que no modelo de RNA a tendência é de subestimar os preços das habitações com valores mais baixos.

No caso do MPH a situação é diferente, isto porque, a diferença entre as percentagens do erro positivo (58,43%) e negativo (41,57) é de 16,86%. Onde o erro positivo médio é de 1.519.239,33 escudos, equivalente a 13.811,26 euros, e o erro negativo médio é de -1.296.410,00 escudos, equivalente a -11.785,55 euros. Isto significa que, o MPH a tendência é superestimar o preço das habitações com valores mais altos e de subestimar o preço das habitações com valores mais baixo, como mostra a figura 5.

Para melhor comparação dos modelos calculamos três medidas erro, o erro quadrado médio (MSE), erro absoluto médio (MAE) e raiz quadrada do erro quadrado médio (RMSE), essas medidas foram escolhidos com base ao estudo (Selim, 2009). Na Tabela 16 estão previstos os resultados de todas as medidas utilizadas no estudo.

Tabela 16: **Modelo de RNA Vs. MPH**

	<b>MSE</b>	<b>RMSE</b>	<b>MAD</b>
<b>RNA</b>	107.190.760,00	10.353,29	6.718,65
<b>MPH</b>	1.805.614.380,68	42.492,52	1.637

Nota: M=40; conjunto de teste. Onde MSE é o erro quadrado médio, MAE é o erro absoluto médio, RMSE é a raiz quadrada do erro quadrado médio.

Podemos verificar através da análise das diferentes medidas de erro que a estimativa de preços com o processo de rede neural é melhor do que a metodologia de estimativa de preços hedônicos. Nota-se que o erro do MPH é sempre superior ao erro do modelo RNA, com uma percentagem superior a 70%.

Uma outra medida de desempenho utilizada pelo SPSS para ajustar a rede neuronal é o erro relativo. A partir desse erro podemos calcular a eficiência da rede neuronal. Esta eficiência é calculada com base no conjunto de teste (Inmaculada Pulido-Calvoa *et al.*, 2007).

Para calcular o erro relativo utilizamos os mesmos conjuntos de testes usados para calcular as outras medidas de desempenho. Aplicando agora as mesmas fórmulas para o modelo do preço hedônicos, obtivemos resultados piores que corrobora a superioridade dos resultados obtidos com redes neurais como mostra a Tabela 17. Esta conclusão também foi obtida no estudo de Teixeira (2011).

Tabela 17: **Erro relativo e eficiência dos modelos**

	<b>Erro relativo</b>	<b>Eficiência do modelo</b>
<b>RNA</b>	0,024	97,6%
<b>MPH</b>	0,097	90,3%

Tabela 17: **Erro relativo e eficiência dos modelos** (cont.)

**Nota.** O erro relativo foi calculado baseando-se na seguinte fórmula:

$$ER = \frac{\sum_{m=1}^m (y^m - \hat{y}^m)^2}{\sum_{m=1}^m (y^m - \bar{y}^m)^2} ;$$

A eficiência dos modelos foi calculada baseando-se na seguinte fórmula:

$$ER = 1 - \frac{\sum_{m=1}^m (y^m - \hat{y}^m)^2}{\sum_{m=1}^m (y^m - \bar{y}^m)^2} .$$

Agora vamos realizar a comparação dos modelos utilizando a variável área.

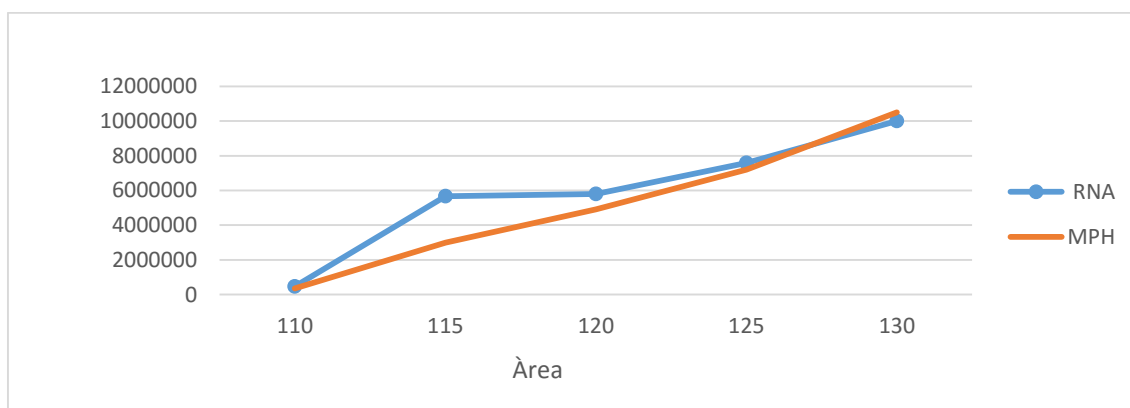


Figura 6. **Comparação dos modelos com a variável área**

Enquanto que o aumento de 5 m<sup>2</sup> de superfície, leva a um aumento aproximadamente de 7% da área das habitações, no modelo de preço hedônico, no caso do modelo de rede neural é diferente, onde mostra claramente a um crescimento desproporcional do gráfico. Entre 110 a 125 m<sup>2</sup> o modelo de redes neuronais artificiais estima o preço das habitações sempre valores superiores do preço estimado pelo modelo de preços hedônico. Entre 125 e 130 m<sup>2</sup>, a estimativa pelos dois modelos é quase a mesma, mais pode-se observar que, num determinado momento o valor estimado entre os dois modelos é igual.

Em suma os resultados mostram que em relação ao modelo MPH, o modelo de RNA é a melhor alternativa para a previsão dos preços das habitações na Ilha do Sal em termos das medidas de desempenho comum analisadas. Verifica-se melhor capacidade de previsão do preço das habitações na ilha no modelo de RNA , uma vez que se verificou os erros entre o preço observado e o preço estimado são inferior ao MPH.

## 5. CONCLUSÃO

O presente estudo teve como objetivo desenvolver um modelo de redes neuronais artificiais para aferir as características das habitações que mais influenciam o preço na Ilha do Sal, em Cabo Verde. Para além das análises baseadas no desenvolvimento do modelo de redes neuronais, afetou-se também a análise pela estimação do modelo de preços hedónicos, onde foram consideradas algumas das características das habitações. Considerou-se uma amostra de 1092 dados no período de 2009 a 2014.

Relativamente aos modelos, considera-se como variável dependente o preço de venda, e considera-se a área, o número de quartos, a existência de varandas, existência de terraços, o número das casas de banho, a localização do imóvel, o número de andares e a proximidade com instituições públicas são como variáveis independentes. Na tentativa de melhorar o desempenho da rede, foi criada a variável zona, onde todas as zonas localizadas em Santa Maria foram agrupadas e denominadas de zona 1, sendo as restantes zonas denominadas de zona 2, esta experiência foi bem-sucedida, uma vez que com a introdução dessa variável o erro relativo de rede diminuiu de 0.485 para 0.476.

Os resultados da experiência com o modelo de RNA e MPH indicaram que as variáveis que têm maior poder explicativo na determinação do preço da habitação na ilha do Sal são a área, a localização e o número de quartos, apesar de que com o MPH algumas variáveis de localização como Bairro Novo, Bairro Novo II, Chã Matias- Espargos, Norte de Santa Maria, Hortelão, Palmeira, Morro curral- espargos e Praia António Sousa- Santa Maria apresentam um efeito negativo no preço. Esta situação pode estar relacionada com o facto de a ilha ser um local turístico. No modelo de RNA a variável perto da câmara municipal e das finanças tem pouco poder explicativo e no MPH a esta variável não apresenta evidências estatisticamente significativa da sua influência no preço das habitações na Ilha do Sal.

Os resultados também mostram que em relação ao modelo MPH, o modelo de RNA é a melhor alternativa para a previsão dos preços das habitações na Ilha. Esta conclusão obteve-se através da comparação dos erros de estimação, dos preços e também em termo das medidas de desempenho comuns analisadas (MSE, MAR e RMSE). É de realçar que o erro do MPH é sempre superior ao erro do modelo RNA.

As principais limitações do trabalho estão essencialmente relacionadas com a dificuldade de obtenção dos dados. Posteriormente, existiram também dificuldades na obtenção de um bom desempenho da rede com os dados disponibilizados. Assim, na tentativa de suprir essas dificuldades foram construídas 6 cenários, onde os resultados indicaram o cenário 6 como o melhor cenário.

Salientar-se que não foi encontrado nenhum estudo com o uso da metodologia de RNA nem do MPH em Cabo Verde, até á data de hoje, considerando-se assim o estudo percussor da temática no País. Para pesquisas futuras sugere-se um estudo desta natureza para outras Ilhas, nomeadamente na Ilha de Santiago, que por ser a maior Ilha possivelmente terá variáveis com maior poder explicativas. Uma das variáveis possíveis de utilização será o estado de conservação das habitações.

## 6. BILIOGRAFIA

- Amaro, J. (2012). *Modelo de avaliação em massa: redes neuronais artificiais aplicada ao sector imobiliário residencial em Portugal. estudo de caso na cidade de lisboa*. Tese de Mestrado Universidade Técnica de Lisboa, Portugal.
- Bao, Q., Wang, Q., & He, L. (2010). A survey for real estate early warning methods. *Anais de 2010 International Conference on Management and Service Science (MASS), Wuhan, China*.
- Bello, A., & Moruf, A. (2010). Does the functional form matter in the estimation of hedonic price model for housing market? *The Social Sciences*, 5(6), 55–564.
- Berry, M., & Linoff, G. (2004). *Data Mining Techniques For Marketing, Sales, and Customer Relationship Management*. (Cap.7, pp.211-257). Indiana: Wiley Publishing.
- Borst, R. (1991). Artificial neural networks: The next modellin/calibration technology for the assessment community? *Property Tax Jornal, International Association of Assessing Officers*, 10(1), 69–94.
- Brooks, C., & Tsolacos, S. (2003). International evidence on the predictability of returns to securitized real estate assets : econometric models versus neural networks International evidence on the predictability of returns to securitized real estate assets : econometric models versus neu. *Journal of Property Research*, 20(2), 133–155.
- Carreira, C. (2011). *The housing market in the district of leiria : a hedonic approach*. Tese de Mestrado, Escola Superior de Tecnologia e Gestão- Instituto Politécnico de Leiria, Portugal.
- Chiarazzoa, V., Caggiania, L., Marinellia, M., & Ottomanellia, M. (2014). A Neural Network based model for real estate price estimation considering environmental quality of property location. *Transportation Research Procedia*, 3(2014), 810–817.
- Collins, B., & Evans, A. (1994). Aircraft noise and residential property values: An artificial neural network approach. *Journal of Transport Economics and Policy*, 28(2), 175–197.

- Cortez, P. (1997). *Algoritmos genéticos e redes neuronais na previsão de séries temporais*. Tese de Mestrado, Universidade do Minho, Portugal.
- Couto, P., Manso, A., & Soeiro, A. (2006). Análise comparativa de valores de mercado de imóveis para habitação. *Anais do 2º Encontro Encontro Nacional sobre Qualidade na Construção*, Lisboa, Portugal.
- Do, A., & Grudnitski, G. (1993). A neural network analysis of the effect of age on housing values. *The Journal of Real Estate Research*, 8(2), 253–264.
- Edmar, M. (1999). *Extração de conhecimento de redes neurais artificiais*. Tese de Mestrado, Instituto de Ciências Matemáticas e de Computação, Universidade De São Paulo, Brasil.
- Fausett, L. (1994). *Fundamental of neural networks: architectures, algorithms and applications*. (Cap. 1, pp. 22-26). Florida: Prentice-Hall, Inc.
- Furtado, B. (2007). Mercado imobiliário e a importância das características locais: uma análise quantílico - espacial de preços hedônicos em Belo Horizonte. *Revista Análise Económica*, 47(9), 71–98.
- Ge, X. J., & Runeson, G. (2004). Modeling property prices Using neural network model for hong kong. *International Real Estate Review*, 7(1), 121–138.
- Gecynalda, S., Gomes, S., & Ludermir, T. B. (2008). Redes neurais artificiais com função de ativação complemento lo-log e probit para aproximar funções na presença de observações extremas. *Revista Da Sociedade Brasileira de Resdes Neurais*, 6(2), 1–12.
- Hamzaoui, Y., Alfredo, J., & Perez, H. (2011). Application of artificial neural networks to predict the selling price in the real estate valuation process. *Anais 2011 International Conference on Artificial Intelligence, Mexican*.

- Igbinsosa, S. (2011). Determinants of residential property value in Nigeria – a neural network approach. *Journal International Multidisciplinary*, 5(5), 152–168.
- Kauko, T. (2003). On current neural network applications involving spatial modelling of property prices. *Journal of Housing and the Built Environment*, 18(2), 159–181.
- Kershaw, P., & Rossini, P. (1999). Using neural networks to estimate constant quality house price indices. *Anais 1999 Fifth Annual Pacific-Rim Real Estate Society Conference, Kuala Lumpur, Malaysia*.
- Kim, K., & Park, J. (2005). Segmentation of the housing market and its determinants: Seoul and its neighbouring new towns in Korea. *Australian Geographer*, 36(2), 221–232.
- Kong, F., Yin, H., & Nakagoshi, N. (2007). Using GIS and landscape metrics in the hedonic price modelling of the amenity value of urban green space: A case study in Jinan City, China. *Landscape and Urban Planning*, 79(3-4), 240–252.
- Lai, P. (2006). Analysis of the mass appraisal model using artificial neural network in Kaohsiung city. *Anais da 23<sup>rd</sup> Pan Pacific Congress of Appraisers, Valuers and Counsellors, San Francisco California*.
- Lam, K., Yu, C., & Lam, K. (2008). An artificial neural network and entropy model for residential property price forecasting in Hong Kong. *Journal of Property Research*, 25(4), 321–342.
- Lima, F., Kimura, H., Neto, A., & Perera, J. (2010). ARIMA-GARCH e redes neurais com ondaletas: velhas tecnologias – novos resultados. *R.Adm., São Paulo*, 45(2), 188–202.
- Limsombunchai, V., Gan, C., & Lee, M. (2004). House price prediction: hedonic price model vs. artificial neural network. *American Journal of Applied Sciences*, 1(3), 193–201.
- Luiz, P., & Soares, B. (2011). Aplicação de redes neurais artificiais em conjunto com o método vetorial da propagação de feixes na análise de um acoplador direcional baseado em fibra ótica. *Revista Brasileira de Computação Aplicada*, 3(2), 58–72.

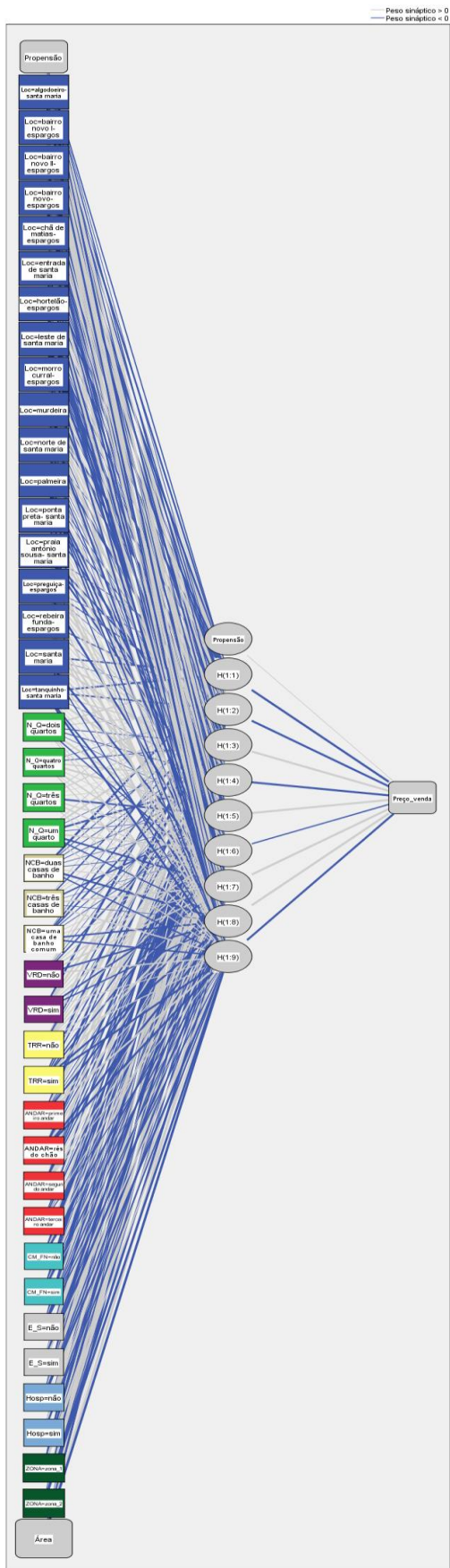
- Mccluskey, W. J., Mccord, M., Davis, P. T., Haran, M., Mcilhatton, D., Mccord, M., Mcilhatton, D. (2013). Prediction accuracy in mass appraisal : a comparison of modern approaches. *Journal of Property Research*, 30(4), 239–265.
- Moreira, D., Silva, S., & Maria, A. (2010). Engenharia de avaliações de imóveis apoiada em técnicas de análise multicritério e redes neurais artificiais. *Revista de Sistemas de Informação Da FSMA*, 6(2), 49–58.
- Juliano, M. (2007). *Um estudo de caso para determinação de preço de vendas de imóveis urbanos via redes neurais artificiais e métodos estatísticos multivariados*. Tese de Mestrado, Universidade Federal do Paraná, Brasil.
- Neto, A. (2006). *Redes neurais artificiais aplicada ás avaliações em massa: estudo de caso para a cidade de Belo Horizonte/MG*. Tese de Mestrado, Univercidade Federal de Minas Gerais.
- Neves, J., Montezuma, J., & Laia, A. (2010). *Análise de Investimentos Imobiliários*. (Cap.1, pp. 22-28).Lisboa: Texto.
- Nguyen, N., & Cripps, A. (2001). Predicting housing value: a comparison of multiple regression analysis and artificial neural networks. *Journal of Real Estate Research*, 22(3), 314–336.
- Pagourtzi, E., Assimakopoulos, V., Hatzichristos, T., & French, N. (2003). Real estate appraisal: a review of valuation methods. *Journal of Property Investment & Finance*, 21(4), 383–401.
- Passari, A. (2003). *Exploração de dados atomizados para previsão de vendas no varejo utilizando redes neurais índice*. Tese de Mestrado, Universidade De São Paulo, Brasil.
- Peddy, L. (2011). Analysis of the mass appraisal model by using artificial neural network in kaohsiung city. *Journal of Modem Accounting and Auditing*, 7(10), 1080–1089.
- Pinto, J. (2013). *Análise comparativa de valores de imóveis e de métodos de avaliação*. Tese de Mestrado, Instituto Politécnico do Porto, Portugal.

- Rossini, P. (1997). Artificial Neural Networks versus Multiple Regression in the Valuation of Residential Property. *Australian Land Economics Review*, 3(1), 1–9.
- Rossini, P. (1998). Improving the results of artificial neural network models for residential valuation. *Anais de 1998 Fourth Annual Pacific-Rim Real Estate Society Conference, Perth, Western Australia*.
- Selim, H. (2009). Determinants of house prices in Turkey: Hedonic regression versus artificial neural network. *Expert Systems with Applications*, 36(2), 2843–2852.
- Selim, S. (2008). Determinants of house prices in Turkey a hedonic regression model. *Dogus Universitesi Dergisi*, 9(1), 65–76.
- Tajani, F., Morano, P., Locurcio, M., & Addabbo, N. (2015). Property valuations in times of crisis. artificial neural networks and evolutionary algorithms in comparison. (Cap.14, pp.194-209). Springer International Publishing.
- Tarré, A. (2009). *Análise de valores de avaliação de apartamento no âmbito do critério a habitação, para duas zonas distintas do conselho de Lisboa- recurso a Modelos Hedónicos*. Tese de Mestrado, Universidade Técnica de Lisboa, Portugal.
- Teixeira, M. (2011). *Modelos econométrico para el precio de los inmuebles: un caso de estudio em portugal*. Tese de Doutoramento Universidad de Córdoba, Espanha.
- Thayer, M., Albers, H., & Rahmatian, M. (1992). The benefits of reducing exposure to waste disposal sites: A hedonic housing value approach. *The Journal of Real Estate Research*, 7(3), 265–282.
- Weirick, W. N., & Ingram, F. J. (1990). Functional form choice in applied real estate analysis. *Appraisal Journal*, 58(1), 57–73.
- Worzala, E., Lenk, M., & Silva, A. (1995). An exploration of neural networks and its application to real estate valuation. *The Journal of Real Estate Research*, 10(2), 185–201.
- Zhang, G., Patuwo, B., & Hu, M. (1998). Forecasting with artificial neural networks: the state of the art. *International Journal of Forecasting*, 14(1), 35–62.
- Zhang, X., Tan, Y., Shen, L., & Wu, Y. (2011). An alternative approach of

competitiveness evaluation for real estate developers. *International Journal of Strategic Property Management*, 15(1), 10–25.

## **7. ANEXO**

**Anexo 1:** – Está a esquema da rede, assim como o peso de sinápticos e os seus respetivos sinais



Fonte: Elaborado pelo SPSS

## Anexo 2: Estão os pesos de conexões e o parâmetro de estimação.

Previsor		Estimativas do parâmetro									
		Previsto									Camada de saída
		Camada oculta 1									Preço_venda
		H(1:1)	H(1:2)	H(1:3)	H(1:4)	H(1:5)	H(1:6)	H(1:7)	H(1:8)	H(1:9)	
Camada de entrada	(Propensão)	-,219	-,060	-,213	,217	-,142	,537	,090	-,509	,327	
	[Loc=algodoeiro- santa maria ]	-,385	-,237	,132	-,340	,461	-,182	,706	,836	-,451	
	[Loc=bairro novo I- espargos ]	,138	-,264	,445	,025	,462	,430	,499	,308	,490	
	[Loc=bairro novo II- espargos ]	-,533	-,421	-,521	-,170	-,344	,037	-,278	,187	,068	
	[Loc=bairro novo- espargos ]	,752	,464	-,214	,442	-,169	-,461	-,063	-,434	-,312	
	[Loc=chá de matias- espargos ]	,349	,223	,236	,107	-,226	-,042	-,249	,327	,035	
	[Loc=entrada de santa maria ]	-,442	-,464	,450	-,121	,152	,129	-,482	,962	-,329	
	[Loc=hortelão- espargos ]	-,185	,372	-,022	,007	-,086	-,122	-,656	-,692	-,389	
	[Loc=leste de santa maria ]	-,338	-,046	-,353	-,494	-,210	-,533	,057	,059	-,398	
	[Loc=morro curral- espargos ]	-,197	,398	-,131	,026	-,503	,385	,062	,036	,223	
	[Loc=murdeira ]	-,960	-,292	,182	,424	-,084	,393	-,111	-,450	,187	
	[Loc=norte de santa maria ]	-,080	,568	-,634	,499	-,026	-,026	-,189	-,023	,190	
	[Loc=palmeira ]	,078	,603	-,106	-,079	,075	-,466	-,418	-,360	-,307	
	[Loc=ponta preta- santa maria ]	,374	-,270	,048	-,253	,082	,248	,510	,040	-,181	
	[Loc=praia antônio sousa- santa maria ]	-,085	,201	-,388	-,462	-,435	,004	,098	,322	,750	
	[Loc=preguiça- espargos ]	-,071	-,240	-,037	,326	,324	,172	-,529	,024	,429	
	[Loc=rebeira funda- espargos ]	,013	-,243	,056	-,025	,103	,020	,465	,377	,282	
	[Loc=santa maria ]	-,436	-,043	,403	,730	,024	,407	-,502	,256	-,208	
	[Loc=tanquinho- santa maria ]	-,466	-,390	,295	-,036	,572	,029	,079	-,416	-,510	
	[N_Q=dois quartos ]	,403	,240	-,160	-,167	-,049	,403	-,262	,125	,205	
	[N_Q=quatro quartos]	,045	-,034	-,171	,350	,539	,090	,388	,077	,266	
	[N_Q=três quartos ]	,112	,224	,033	,075	-,499	,287	,351	,043	,167	
	[N_Q=um quarto ]	,170	,465	-,059	-,546	-,078	,459	-,464	-,280	,188	
	[NCB=duas casas de banho ]	-,709	-,134	-,125	-,103	-,059	-,080	,172	-,532	-,283	
	[NCB=três casas de banho ]	,134	-,267	,073	,191	-,374	-,023	,426	-,057	,348	
	[NCB=uma casa de banho comum ]	-,026	-,312	-,389	-,127	-,084	,109	,152	-,407	-,381	
	[VRD=não]	,054	,341	-,054	-,307	,169	,372	,382	-,450	,346	
	[VRD=sim]	-,651	,002	,472	,336	,074	,360	-,406	-,318	,137	
	[TRR=não]	,417	-,241	-,035	,213	,353	,290	,066	,200	,183	
	[TRR=sim]	-,807	,181	-,253	,023	-,251	-,338	,548	-,577	-,629	
	[ANDAR=primeiro andar]	-,552	-,199	,520	,285	-,072	-,065	-,459	,199	,460	
	[ANDAR=rés do chão ]	,242	,234	-,450	-,125	-,155	,383	,047	-,689	-,744	
	[ANDAR=ssegundo andar ]	,463	-,111	-,289	-,234	-,245	,145	,089	,294	,144	
	[ANDAR=terceiro andar ]	,008	-,293	,010	-,045	-,284	,061	-,013	,178	-,131	
	[CM_FN=não ]	,485	-,251	-,537	-,516	-,178	,071	,259	-,398	-,037	
	[CM_FN=sim ]	-,324	-,527	-,277	-,071	,335	-,098	-,149	-,298	,065	
	[E_S=não]	-,668	-,617	-,129	,406	-,113	-,151	-,267	-,229	-,226	
	[E_S=sim]	,225	-,357	,308	,124	,353	-,393	-,237	-,714	,269	
	[Hosp=não]	-,504	-,691	,290	-,338	,478	,412	-,222	-,304	-,487	
	[Hosp=sim]	-,179	,212	,401	,148	,158	,121	,383	-,222	-,278	
[ZONA=zona_1]	-,786	-,052	,123	-,354	-,387	,109	-,108	,706	,304		
[ZONA=zona_2]	,885	-,100	-,905	,160	,371	,547	-,121	-,1148	-,665		
Área	,048	-,1863	,924	1,374	-,701	,046	,025	,807	-,1,394		
Camada oculta 1	(Propensão)										,079
	H(1:1)										-,1,082
	H(1:2)										-,719
	H(1:3)										,706
	H(1:4)										-,720
	H(1:5)										,608
	H(1:6)										-,276
	H(1:7)										,797
	H(1:8)										,899
H(1:9)										-,691	

Fonte: Elaborado pelo SPSS